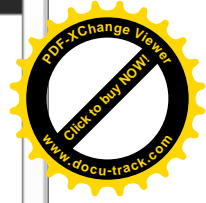
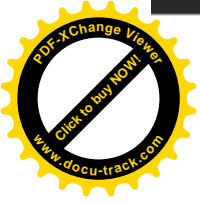


H. J. GODWIN, M.A.

INEQUALITIES ON  
DISTRIBUTION  
FUNCTIONS

BEING  
NUMBER SIXTEEN  
OF

GRIFFIN'S STATISTICAL  
MONOGRAPHS & COURSES  
EDITED BY M. G. KENDALL, M.A., Sc.D.



# INEQUALITIES ON DISTRIBUTION FUNCTIONS

H. J. GODWIN, M.A.

*Senior Lecturer, Department of Pure Mathematics  
University College of Swansea*

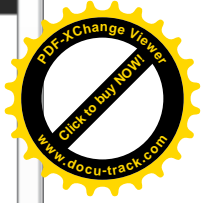
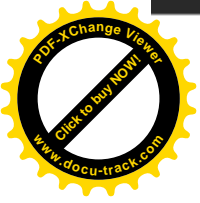
---

BEING NUMBER SIXTEEN OF  
GRIFFIN'S STATISTICAL  
MONOGRAPHS & COURSES

EDITED BY  
M. G. KENDALL, M.A., Sc.D.

---

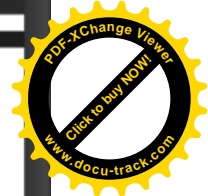
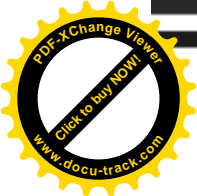
CHARLES GRIFFIN & COMPANY LIMITED  
LONDON



*Copyright © 1964*  
**CHARLES GRIFFIN & COMPANY LIMITED**  
**42 DRURY LANE, LONDON, W.C.2**

**First published in 1964**

**Printed in Great Britain by Chorley & Pickersgill Ltd**



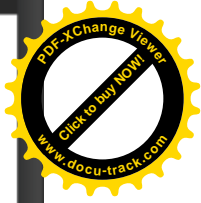
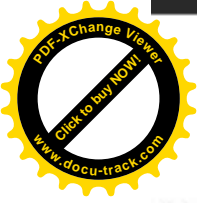
## PUBLISHERS' NOTE

The series in which this title appears was introduced by the publishers in 1957 and is under the general editorship of Dr. Maurice G. Kendall. It is intended to fill a need which has been evident for some time and is likely to grow — the need for some form of publication at moderate cost which will make accessible to a group of readers specialized studies in statistics or special courses on particular statistical topics. There are numerous cases where, for example, a monograph on some newly developed field would be very useful, but the subject has not reached the stage where a comprehensive book is possible; or, again, where a course of study is desired in a domain not covered by textbooks but where an exhaustive treatment, even if possible, would be expensive and perhaps too elaborate for the readers' needs.

Considerable attention has been given to the problem of producing these books speedily and economically. Appearing in a cover the design of which will be standard, the contents of each volume will follow a simple, straightforward layout, the text production method adopted being suited to the complexity or otherwise of the subject.

The publishers will be interested in approaches from any authors who have work of importance suitable for the series.

CHARLES GRIFFIN & CO. LTD.



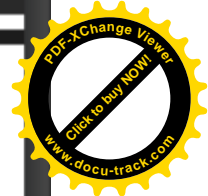
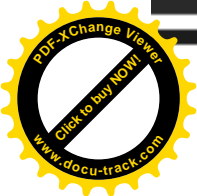
## OTHER BOOKS ON MATHEMATICS AND STATISTICS

<i>Games, gods and gambling</i>	F. N. DAVID
<i>Combinatorial chance</i>	F. N. DAVID and D. E. BARTON
<i>A statistical primer</i>	F. N. DAVID
<i>An introduction to the theory of statistics</i>	G. U. YULE and M. G. KENDALL
<i>The advanced theory of statistics (three volumes)</i>	M. G. KENDALL and A. STUART
<i>Rank correlation methods</i>	M. G. KENDALL
<i>Exercises in theoretical statistics</i>	M. G. KENDALL
<i>Rapid statistical calculations</i>	M. H. QUENOUILLE
<i>The design and analysis of experiment</i>	M. H. QUENOUILLE
<i>Sampling methods for censuses and surveys</i>	F. YATES
<i>Biomathematics</i>	C. A. B. SMITH
<i>Statistical method in biological assay</i>	D. J. FINNEY
<i>The mathematical theory of epidemics</i>	N. T. J. BAILEY
<i>Probability and the weighing of evidence</i>	I. J. GOOD

### GRIFFIN'S STATISTICAL MONOGRAPHS AND COURSES:

No. 1: <i>The analysis of multiple time-series</i>	M. H. QUENOUILLE
No. 2: <i>A course in multivariate analysis</i>	M. G. KENDALL
No. 3: <i>The fundamentals of statistical reasoning</i>	M. H. QUENOUILLE
No. 4: <i>Basic ideas of scientific sampling</i>	A. STUART
No. 5: <i>Characteristic functions</i>	E. LUKACS
No. 6: <i>An introduction to infinitely many variates</i>	E. A. ROBINSON
No. 7: <i>Mathematical methods in the theory of queueing</i>	A. Y. KHINTCHINE
No. 8: <i>A course in the geometry of <math>n</math> dimensions</i>	M. G. KENDALL
No. 9: <i>Random wavelets and cybernetic systems</i>	E. A. ROBINSON
No. 10: <i>Geometrical probability</i>	M. G. KENDALL and P. A. P. MORAN
No. 11: <i>An introduction to symbolic programming</i>	P. WEGNER
No. 12: <i>The method of paired comparisons</i>	H. A. DAVID
No. 13: <i>Statistical assessment of the life characteristic: a bibliographic guide</i>	W. R. BUCKLAND
No. 14: <i>Applications of characteristic functions</i>	E. LUKACS and R. G. LAHA
No. 15: <i>Elements of linear programming: with economic applications</i>	R. C. GEARY and M. D. MCCARTHY

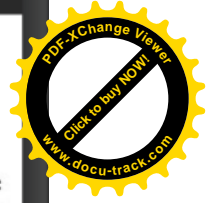
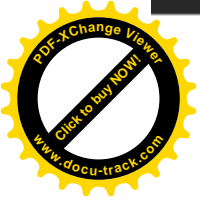
Descriptive leaflets available from Charles Griffin & Co. Ltd.



## PREFACE

The essence of mathematics is to take a given set of facts and to deduce their consequences. In the problems discussed in this monograph the initial facts are properties of the distributions of random variables, and the consequences which are of interest are the bounds which may be placed on the probability of the variables taking values belonging to some given set. The extreme situations are the one in which the distributions are completely determined and the probability is also completely determined, and that in which nothing is known about the distributions and the probability may be any number between 0 and 1 inclusive. Between these extremes lie the cases in which, from some knowledge of the distributions, we can say something which is not trivial about the probability. The known facts about a distribution may be numerical, e.g. that it has certain moments taking certain values, or geometrical, e.g. that it has a single mode or that the graph of its probability density function is smooth according to some criterion. The type of fact which is taken as known and the type of set considered are indicated in the chapter and section headings. Page references in the bibliography act as an index of names.

The concept of convexity is found to be a valuable way of unifying much of the work, and in Chapter I an account is given of the ideas and results which are needed subsequently. In Chapter II we deal with the univariate distributions for which the data are expectations of various functions. When these functions do not depend on the distribution function it is possible to give a complete solution of the problem (though one which, in practice, may still involve some complicated calculation). The case of the mean range, which is expressible as the expectation of a function of the distribution function, is considered as a special case in Section 2.10. In Chapter III we consider univariate distributions for which geometrical data are given. The method used in Chapter II can be



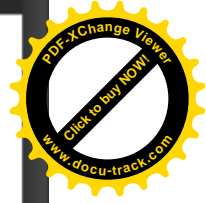
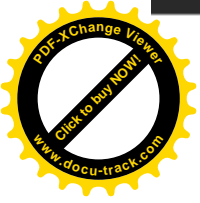
adapted for many cases here also, though it fails in the case of the range or when we are given restrictions on the magnitude of functions. In Chapter IV we deal with multivariate distributions; we confine ourselves to the case when the data are second-order moments, and even with this simplification it appears that the computational difficulties involved in finding a best possible solution are formidable. In Chapter V we consider not single variables, but sums of variables; here no general methods seem available and various problems are treated by *ad hoc* methods which yield results far from best possible. Finally, in Chapter VI there are some notes on the applicability of the results obtained in earlier chapters from the point of view of their value as sources of problems to the pure mathematician and for application by the statistician. The reader may find it useful to look at this chapter before going on to the detailed ones which precede it.

Throughout the monograph the emphasis is on methods and solutions which lead to definite numerical bounds. Some recent work has been concerned with generalizing the earlier ideas, but where these generalizations do not give rise to concrete results they are mentioned only briefly.

At the end of the monograph is a set of exercises; it seems convenient to refer in these to work which, with hindsight, we may regard as elementary. It is not to be considered that the efforts of pioneers are being disparaged by being treated in this way.

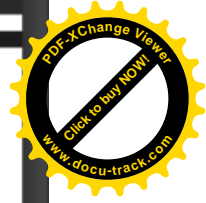
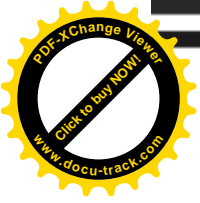
1963

H. J. GODWIN



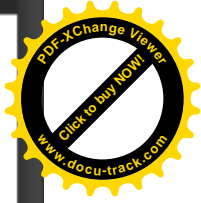
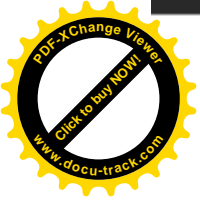
## CONTENTS

	<i>Page</i>
<b>I PRELIMINARIES</b>	
1.1 Notation	9
1.2 Convexity	10
<b>II UNIVARIATE DISTRIBUTIONS: NUMERICAL DATA</b>	
2.1 Introduction	13
2.2 Properties of moments	13
2.3 The application of convexity	16
2.4 A numerical example	18
2.5 Single interval; first and second moments given	21
2.6 Half line; ordinary moments given	22
2.7 Absolute moments given	26
2.8 Finite interval	28
2.9 Expectations of general functions given	29
2.10 Mean range given	32
<b>III UNIVARIATE DISTRIBUTIONS: GEOMETRICAL DATA</b>	
3.1 Introduction	37
3.2 Unimodal distribution: second moment about mode given	37
3.3 Unimodal distribution: first and second absolute moments about mode given	39
3.4 Unimodal distribution: first and second moments about any point given	43
3.5 A numerical example	45
3.6 Restrictions on the magnitude of $f(x)$ or $F(x)$	47
3.7 Mean range for unimodal symmetrical distribution	52
<b>IV MULTIVARIATE DISTRIBUTIONS</b>	
4.1 Introduction	56
4.2 Second-order moments: rectangular region	56
4.3 Second-order moments: region similar to an orthant	61
4.4 A continuous set of variables	62
4.5 Variances given: rectangular regions	63
4.6 Bivariate distribution: convex polygonal region	64
4.7 A general theorem (first- and second-order moments given)	66
4.8 Elliptical region: independent variables	69
4.9 A monotonicity condition: ellipsoidal region	72



## CONTENTS

	<i>Page</i>
<b>V SUMS OF VARIABLES</b>	
5.1 Introduction	77
5.2 Population and sample variances given	77
5.3 Symmetrical unimodal bounded distribution	79
5.4 Comparison with normal distribution	81
5.5 Restrictions on all moments	82
5.6 An inequality for partial sums	83
<b>VI APPLICATIONS</b>	86
<b>EXERCISES</b>	89
<b>BIBLIOGRAPHY</b>	94



## CHAPTER I

### PRELIMINARIES

#### 1.1 Notation

Since there is considerable variation among different writers in the matter of notation, we summarize here the notation which we shall use.

$(x_1, \dots, x_n)$  denotes a random variable from an  $n$ -dimensional population; when  $n = 1$  we drop the suffix.

$F(x_1, \dots, x_n)$  is the distribution function of the population, i.e.  $F(X_1, \dots, X_n)$  is the probability that  $x_1 \leq X_1, \dots, x_n \leq X_n$ . If the  $x_i$  take discrete values then  $F$  is a step-function: if the  $x_i$  take continuous sets of values then  $(\partial^n F)/(\partial x_1 \dots \partial x_n)$  is called the p.d.f. (probability density function) and is denoted by  $f(x_1, \dots, x_n)$ . It may be noted in passing that the distinction between discrete and continuous distributions is of little importance in the present work since a distribution of either type can be approximated to arbitrarily closely by one of the other type, and so we obtain no improvement in inequalities by confining ourselves to one type or the other.

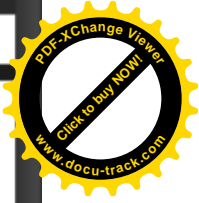
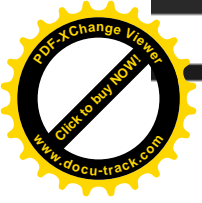
The expectation of the function  $\phi(x_1, \dots, x_n)$ , written  $E(\phi(x_1, \dots, x_n))$ , is

$$\int_{x_1=-\infty}^{\infty} \dots \int_{x_n=-\infty}^{\infty} \phi(x_1, \dots, x_n) dF(x_1, \dots, x_n)$$

where, to cover both the discrete and continuous cases, we interpret the integral in the sense of Stieltjes. If  $f(x_1, \dots, x_n)$  exists we can write the integral as

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \phi(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

For  $n \geq 2$  we shall be concerned only with second-order moments; unless the contrary is specifically stated we shall assume that  $E(x_i) = 0$  ( $i = 1, \dots, n$ ) and we shall write  $E(x_i^2) = \sigma_i^2$ ,  $E(x_i x_j) = \rho_{ij} \sigma_i \sigma_j$ .



For  $n=1$  we denote  $E((x-a)^r)$  by  $\mu_r(a)$ .  $\mu_r(0)$  is written  $\mu_r'$ , and  $\mu_r(\mu_1')$  is written simply  $\mu_r$ . These are ordinary moments; absolute moments  $E(|x-a|^r)$  are denoted by  $\nu_r(a)$ . We do not count  $\mu_0$  or  $\nu_0$ , which are both identically unity, among the moments; by the first  $r$  moments we mean  $\mu_1, \dots, \mu_r$ .  $\nu_r(0)$  is denoted simply by  $\nu_r$ ; note that this convention is different for ordinary moments and for absolute moments — in the former case the absence of the brackets and prime means that we are taking moments about the mean of the population, but in the latter case we are taking them about the origin. Note also that the absolute moments  $\nu_r(a)$  for the population with p.d.f.  $f(x)$  are the same as the ordinary moments  $\mu_r(a)$  for the population with p.d.f. 0 for  $x < a$ ,  $f(x) + f(2a-x)$  for  $a \leq x$ , so that we could express any results in terms of absolute moments in terms of ordinary moments for a population of this kind. (To put the matter in geometrical terms, we can fold<sup>(\*)</sup> the distribution about the line  $x=a$ , and measurement to the right of  $a$  is equivalent to measurement in both directions from  $a$  for the unfolded distribution.)

We shall be concerned with the problem of finding bounds for  $P(T)$ , i.e. the probability that  $(x_1, \dots, x_n)$  lies in a certain set  $T$ ; we denote by  $L(T)$  and  $U(T)$  the infimum and supremum respectively of  $P(T)$ , under a given set of conditions on the distribution. A statement of the form “ $U(T)$  equals something” means that a best possible upper bound has been found for the probability; otherwise we may have to be content with a statement of the form “ $U(T)$  is less than or equal to something”.

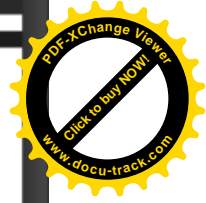
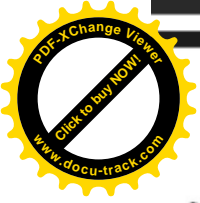
## 1.2 Convexity

A considerable unification of the results which follow can be achieved by the use of the ideas of the theory of convex sets. In this section we give an account of only those propositions which we shall need subsequently, leaving it to the interested reader to pursue the subject further in books such as that by Eggleston (1958).

One definition of a convex set in  $n$ -dimensional Euclidean space is that if the points  $\mathbf{x} (x_1, \dots, x_n)$  and  $\mathbf{y} (y_1, \dots, y_n)$  lie in the set, then

---

(\*) Peek (1933) used this terminology.



so do all the points  $t\mathbf{x} + (1 - t)\mathbf{y}$  for  $0 \leq t \leq 1$ . In geometrical terms, if the set contains the end-points of a straight line of finite length then it contains all its intermediate points.

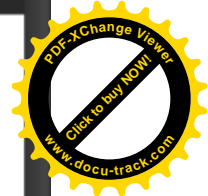
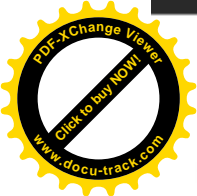
Two convex sets with no point of one in the interior of the other can be separated by a hyperplane, i.e. there exists a linear form  $a_1 x_1 + \dots + a_n x_n$  which is non-negative at all points of one set and non-positive at all points of the other. If neither the sets nor their frontiers have points in common then there is a hyperplane which separates them strictly, so that  $a_1 x_1 + \dots + a_n x_n$  is positive for all points in one set and negative for all points in the other.

A hyperplane which contains at least one point of the frontier of the convex set  $S$  but is such that there is no point of  $S$  in one of the open half-spaces separated by the hyperplane is a support plane of  $S$ . For a bounded set (i.e. one such that the coordinates of all points of the set are bounded) there exist two support hyperplanes in every direction, but this need not be so for an unbounded set (e.g. a half-space).

In the next chapter we shall be interested in a special case which can be stated as follows. The convex set  $S$  which we consider is the union of the half-spaces

$$l(\theta)(\mathbf{x}) \geq 0$$

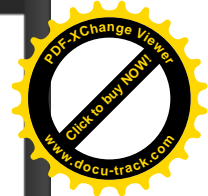
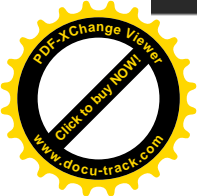
where  $l(\theta)(\mathbf{x}) = l_1(\theta)x_1 + \dots + l_n(\theta)x_n$ .  $\theta$  is some index which takes a set of values which need not be enumerable nor even one-dimensional. We are interested in the values of the linear form  $\psi(\mathbf{x}) = A_1 x_1 + \dots + A_n x_n$  which is known to be non-negative at all points of  $S$  and to take the value one at a point of  $S$ . If  $\psi_0$  is the infimum of  $\psi$  in  $S$  then  $\psi = \psi_0$  at at least one point  $P$  of  $S$  (since  $S$  is closed), and  $P$  cannot lie in the interior of  $S$  since we could decrease  $\psi$  by moving from  $P$  in a suitable direction while still remaining in  $S$ . Since no point at which  $\psi < \psi_0$  lies in  $S$ , the hyperplane  $\psi = \psi_0$  is a support hyperplane of  $S$ . Now let  $d(\theta)$  be the distance of  $P$  from the hyperplane  $l(\theta)(\mathbf{x}) = 0$ ; if  $\inf d(\theta) > 0$  then  $P$  must lie in the interior of  $S$ . Hence  $\inf d(\theta) = 0$ , and if we add to the set  $\mathcal{A}$  of hyperplanes  $l(\theta)(\mathbf{x}) = 0$  the limiting cases of convergent sequences of these hyperplanes (i.e. if we take



the closure  $\overline{A}$  of the set  $A$ ) then  $P$  must lie on at least one hyperplane of  $\overline{A}$ . Since any  $n + 1$  hyperplanes through a point in  $n$ -dimensional space are linearly dependent we can express  $\psi(\mathbf{x}) - \psi_0$  in the form

$$\sum_{i=1}^k \lambda_i l(\theta_i)(\mathbf{x}),$$

where  $k \leq n$  and the summation may include hyperplanes from  $\overline{A}$ . Now if  $\lambda_1 < 0$  we can find a point  $\mathbf{y}$  for which  $l(\theta_1)(\mathbf{y}) > 0$ ,  $l(\theta_2)(\mathbf{y}) = \dots = l(\theta_k)(\mathbf{y}) = 0$ , so that  $\mathbf{y}$  lies in  $S$  and yet  $\psi(\mathbf{y}) < \psi_0$ . This is impossible and so  $\lambda_1 \geq 0$ , and similarly  $\lambda_2, \dots, \lambda_k$  are all non-negative.



## CHAPTER II

# UNIVARIATE DISTRIBUTIONS: NUMERICAL DATA

### 2.1 Introduction

In this chapter we deal with the most extensively studied case and shall show that an almost complete solution is possible, though one which may still present formidable computational difficulties. This is when the data are moments and  $T$  is one or more intervals (possibly extending to  $+\infty$  or  $-\infty$ ). We first note the restrictions which must be placed on a set of numbers if they are to arise as moments of a distribution and then show how, if a set satisfying the restrictions is given, we can obtain the quantities  $U$  and  $L$ . Some algebraic solutions for simple forms of  $T$  are then obtained. We next consider the case when we are given expectations of more general functions than powers of the variable. Finally we consider the case of the mean range, which can be represented as the expectation of a function of the distribution function and so is of quite a different type from the other expected values which are used.

### 2.2 Properties of moments

Since a distribution function is required to be non-decreasing, it is not possible to assign arbitrary values to a set of moments and then find a distribution which will give these values. The consideration of the relations which must exist between moments was one which occupied analysts for many years, and in this monograph we shall discuss only a few points which have relevance to what follows. For further information on the subject we refer the reader to Shohat and Tamarkin (1943).

Suppose that  $\mu'_1, \dots, \mu'_{2n}$  are given for a distribution; if we replace  $f(x)$  by  $(1 - \epsilon^{2n+1})f(x)$  with additional probability  $\epsilon^{2n+1}$  at  $a/\epsilon$  then we obtain a distribution with moments  $\mu_r^*$  where

$$\mu_r^* = \mu_r' + 0(\epsilon) \quad \text{for } r \leq 2n$$

$$\mu_{2n+1}^* = \mu_{2n+1}' + a^{2n+1}$$

$$|\mu_r^* - \mu_r'| \rightarrow \infty \text{ as } \epsilon \rightarrow 0 \text{ for } r \geq 2n + 2,$$

provided that  $a \neq 0$ . By considering what happens as  $\epsilon$  tends to zero, we see that  $\mu_{2n+1}'$  is independent of the choice of  $\mu_1', \dots, \mu_{2n}'$ . We shall call the process of changing a distribution in this way "adding zero probability at infinity" and shall use it in later work. If we proceed similarly starting with  $\mu_1', \dots, \mu_{2n-1}'$  then we add  $a^{2n}$  to  $\mu_{2n}'$ , and so  $\mu_{2n}'$  can take arbitrarily large values for given  $\mu_1', \dots, \mu_{2n-1}'$  but not arbitrarily small ones (it cannot, of course, be negative). By consideration of the fact that

$$\int_{-\infty}^{\infty} (b_0 + b_1 x + \dots + b_n x^n)^2 dF(x) \quad (2.2.1)$$

is a non-negative form in  $b_0, \dots, b_n$ , we have (see, e.g. Mirsky (1955), p. 400, for the relevant algebraic theorem) that

$$\begin{vmatrix} 1 & \mu_1' & \dots & \mu_n' \\ \mu_1' & \mu_2' & \dots & \mu_{n+1}' \\ & & \dots & \\ \mu_n' & \mu_{n+1}' & \dots & \mu_{2n}' \end{vmatrix} \geq 0. \quad (2.2.2)$$

If the determinant on the left-hand side of (2.2.2) is zero then the form in (2.2.1) is semi-definite and can be zero without vanishing identically; this means that  $b_0, \dots, b_n$  (not all zero) exist such that

$$\int_{-\infty}^{\infty} (b_0 + b_1 x + \dots + b_n x^n)^2 dF(x) = 0.$$

This can be so only if  $dF(x)$  is zero except when  $x$  is a zero of the polynomial  $b_0 + b_1 x + \dots + b_n x^n$  so that the distribution consists of a finite number of discrete probabilities. If the number of points at which  $dF(x)$  is non-zero is less than  $n$ , the distribution is said to be degenerate.

We shall be interested later in using moments to estimate probabilities, and to that extent we may ask whether the above statements have converses; in other words, whether a set of moments

satisfying (2.2.2) determines a distribution. The answer is that at least one distribution is determined but that there may be no unique solution. In order to ensure uniqueness we require further conditions such as that the series

$$\sum_{n=1}^{\infty} \mu_{2n}^{-\frac{1}{2n}}$$

diverge (see Shohat and Tamarkin (1943), p. 20). This means roughly that the moments must not be too large or that the distribution must not be too spread-out. (If the distribution is contained in a finite interval such that  $f(x) = 0$  for  $|x| > A$  then  $\mu_{2n} \leq (2A)^{2n}$  and

$$\sum_{n=1}^{\infty} \mu_{2n}^{-\frac{1}{2n}}$$

diverges; for the normal distribution with unit variance we have  $\mu_{2n} = (2n-1) \dots 3 \cdot 1 < 2^n n!$  and again

$$\sum_{n=1}^{\infty} \mu_{2n}^{-\frac{1}{2n}}$$

diverges.) If the distribution is widely spread-out we can add to  $f(x)$  a multiple of a function such as  $\exp(-|x|^{\frac{1}{2}}) \cdot \cos(|x|^{\frac{1}{2}})$ , all of whose moments are zero, and still obtain a non-negative function which may be taken to be a p.d.f. For example, if

$$f_1(x) = \frac{1}{4} \exp(-|x|^{\frac{1}{2}})$$

$$f_2(x) = \frac{1}{4} \exp(-|x|^{\frac{1}{2}}) (1 + \cos(|x|^{\frac{1}{2}}))$$

then all moments of  $f_1(x)$  are equal to the corresponding moments of  $f_2(x)$  (see Kendall and Stuart (1958, 1963), p. 109), but

$$\int_{-\pi^{1/4}}^{\pi^{1/4}} f_1(x) dx = .4656 \dots$$

$$\int_{-\pi^{1/4}}^{\pi^{1/4}} f_2(x) dx = .7328 \dots$$

Using moments alone, we could never approximate to either probability with an error less than .2672 ....

### 2.3 The application of convexity

Suppose that we are given the expected values  $H_1, \dots, H_k$  of the functions  $h_1(x), \dots, h_k(x)$  and that these expected values can actually be obtained with some distribution.

We consider the linear form

$$g(x) = a_0 + a_1 h_1(x) + \dots + a_k h_k(x)$$

and the values of  $a_0, \dots, a_k$  for which

$$g(x) \geq \chi_T(x), \quad (2.3.1)$$

where  $\chi_T(x)$  is 1 if  $x$  belongs to  $T$  and 0 otherwise. We call  $\chi_T(x)$  the characteristic function of  $T$ .

The set of points  $\mathbf{a}(a_0, \dots, a_k)$  is a convex set  $A$  of the type discussed in Section 1.2. The closure of the set of hyperplanes defining  $A$  is obtained by replacing  $\chi_T(x)$  by  $\chi_T^*(x)$ , where  $\chi_T^*(x) = \max\{\chi_T(x+0), \chi_T(x-0)\}$  and is the characteristic function of the set  $T^*$  formed from  $T$  by including the end-points of intervals in  $T$ . We shall have a different value of  $U$  (or  $L$ ) for  $T^*$  from the value we have for  $T$  only if there is non-zero probability at a point which is in  $T^*$  but not in  $T$ , and this is so only when we have equality in some inequality such as (2.2.2), so that the distribution is necessarily discrete. Thus, although it is  $T^*$  for which we are actually going to evaluate  $U$  and  $L$ , the results will normally be applicable to  $T$  also.

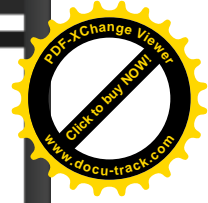
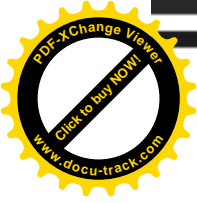
If

$$\psi = \psi(a_0, \dots, a_k) = E(g(x)) = a_0 + a_1 H_1 + \dots + a_k H_k$$

then  $\psi$  is non-negative in  $A$ , and at the point  $(1, 0, \dots, 0)$  of  $A$  we have  $\psi = 1$ . Hence if  $\psi_0$  is the infimum of  $\psi$  over  $A$  then we can express  $\psi - \psi_0$  in the form

$$\sum_1^r \lambda_i (a_0 + a_1 h_1(x_i) + \dots + a_k h_k(x_i) - \chi_T^*(x_i))$$

where  $r \leq k+1$  and the  $\lambda_i$  are non-negative.



On comparing coefficients of  $a_0, \dots, a_k$  we have

$$1 = \sum_1^r \lambda_i$$

$$\sum_1^r \lambda_i h_s(x_i) = H_s \quad (s = 1, \dots, k)$$

$$\psi_0 = \sum_1^r \lambda_i \chi_T^*(x_i).$$

The discrete distribution with probabilities  $\lambda_i$  at  $x_i$  ( $i=1, \dots, r$ ) thus gives the correct expected values and also gives

$$P(T^*) = \sum_1^r \lambda_i \chi_T^*(x_i) = \psi_0$$

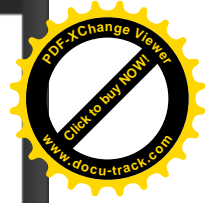
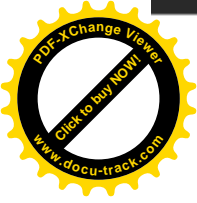
so that  $\psi_0 \leq U$ .

But for any distribution giving the expected values  $H_1, \dots, H_k$  we have  $P(T^*) = E(\chi_T^*(x)) \leq E(g(x)) = \psi$ . This is true for any choice of  $a_0, \dots, a_k$  satisfying (2.3.1) and so  $P(T^*) \leq \psi_0$  and  $U \leq \psi_0$ .

Hence  $U = \psi_0$ .

Similarly we can find  $L$  by taking the supremum of  $\psi$ , subject to the conditions  $g(x) \leq \chi_T(x)$ .

We have thus reduced the problem of finding  $U$  to that of finding the curve which is "lowest" (in the sense of representing the function giving the least expectation) and which lies above a succession of lines at height zero or one above the  $x$ -axis. In the examples which follow we shall show how curves of this kind may be found: for the moment we note that the distributions which we find to give  $U$  (or  $L$ ) are discrete ones, with probability only at the points where  $g = \chi$ . If, by intelligent guesswork, we can construct a discrete distribution which gives the required expected values and for which there exist  $a_0, \dots, a_k$  such that (2.3.1) is satisfied, then we have constructed a support hyperplane to the set  $A$  which is in the correct direction and so is the unique one. Thus our guess may be established as correct without examining all other cases. This possibility of showing that a trial solution is the correct one arises in other applications of convexity such as linear programming; see, for example, Gale (1960), p. 22.



The use of the form  $g(x)$  is due to Isii (1959a); the use of convexity has been discussed by Marshall and Olkin (1960b) in the context of multivariate distributions and by Kingman (1963) who states the problem in a very general form. Isii, for the case when the  $h_i(x)$  are  $x, x^2, \dots, x^{2n}$ , also deals with the more difficult question of whether or not a distribution will actually exist, and shows that under fairly general conditions this is so, provided the moments satisfy the necessary conditions mentioned in Section 2.2.

## 2.4 A numerical example

As an illustration of the method we consider the case in which  $T$  is the pair of intervals  $\frac{1}{2} < |x| < 2$  and we are given  $\mu'_1 = 0, \mu'_2 = 1, \mu'_4 = 3$ . Guttman (1948a) has given a formula which covers certain cases of this type, with two intervals symmetrical about the mean and  $\mu'_3$  unknown, but the example which we are considering now does not fit into his scheme.

We note that

$$\begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & \mu'_3 \\ 1 & \mu'_3 & 3 \end{vmatrix} = 2 - \mu'^2_3$$

is positive for a range of values of  $\mu'_3$ , and so we can assume that the distribution is not discrete, and we can include the end-points  $x = \pm \frac{1}{2}$  and  $x = \pm 2$  in  $T$  or exclude them as desired. If  $\mu'_4 = 1$  then we must have  $\mu'_3 = 0$ , and the only possible distribution has probability  $\frac{1}{2}$  at  $x = -1$  and  $x = 1$ . With the given value of  $\mu'_4$  we can have this distribution together with zero probability at infinity, and so there is a distribution with unit probability in  $T$ . Hence  $U = 1$ .

In order to find  $L$  we consider polynomials  $g(x)$  of the form  $a_0 + a_1x + a_2x^2 + a_4x^4$  satisfying  $g(x) \leq \chi_T(x)$ , i.e.  $g(x) \leq 0$  for  $|x| \leq \frac{1}{2}$  and for  $2 \leq |x|$ , and  $g(x) \leq 1$  for  $\frac{1}{2} < |x| < 2$  (remembering that in finding  $L$  the end-points of the intervals are taken with the lower value of  $\chi_T(x)$ ; in finding  $U$  they would be taken with the higher value).

We suppose for the moment that  $a_4 \neq 0$ .

We want to maximize  $\psi = E(g(x))$  and we certainly increase  $\psi$  if we increase  $g(x)$  for all  $x$ . By increasing  $a_0$  we may suppose that  $g(x) = \chi_T(x)$  at some value  $x = \alpha$ , and then by adding a positive multiple of  $(x - \alpha)^2$  that  $g(x) = \chi_T(x)$  for a second value  $x = \beta$ . If there were a term in  $x^3$  in  $g(x)$  we could also add a positive multiple of  $(x - \alpha)^2(x - \beta)^2$  to get a third value at which  $g(x) = \chi_T(x)$ , but this is not possible with the given conditions. Unless  $L = 0$  we shall have  $\chi_T(x) = 1$  as one of the values at which  $g(x) = \chi_T(x)$ ; we may suppose that  $\alpha$  is such a value and, by symmetry, that  $\frac{1}{2} < \alpha < 2$ . Since the coefficient of  $x^3$  in  $g(x)$  is zero and  $a_4 \neq 0$  the sum of the roots of  $g(x) = \text{constant}$  is zero, and so it would not be possible for  $g(x)$  to have a maximum value 0 at  $\beta$  with  $\beta < -2$  or  $|\beta| < \frac{1}{2}$ ; we may suppose that  $\beta$  is one of  $-2, -\frac{1}{2}, \frac{1}{2}$  or  $2$  with  $g(\beta) = 0$ , or else that  $g(x)$  has a maximum value 1 at  $\beta$  with  $\beta = -\alpha$  (from consideration of the sum of the roots of  $g(x) = 1$ ). Taking this last possibility, we have

$$g(x) = 1 - k(x - \alpha)^2(x + \alpha)^2$$

where  $k$  is some positive number, and we require  $g(2)$  and  $g(-2)$  to be non-positive,

$$\text{i.e.} \quad \sqrt{k}(4 - \alpha^2) \geq 1 \quad (2.4.1)$$

and  $g(\frac{1}{2})$  and  $g(-\frac{1}{2})$  to be non-positive,

$$\text{i.e.} \quad \sqrt{k}(\alpha^2 - \frac{1}{4}) \geq 1. \quad (2.4.2)$$

We have to maximize

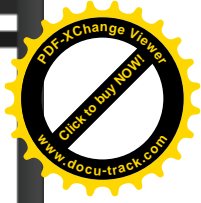
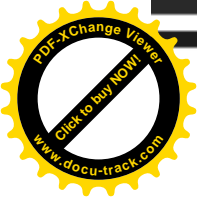
$$\psi = 1 - k\alpha^4 + 2k\alpha^2 - 3k,$$

subject to (2.4.1) and (2.4.2). The bounds for  $k$  given by (2.4.1) and (2.4.2) are equal for  $\alpha^2 = 17/8$ ; hence we have

$$\sqrt{k} \geq (4 - \alpha^2)^{-1} \quad \text{for} \quad 17/8 \leq \alpha^2 < 4$$

and

$$\sqrt{k} \geq (\alpha^2 - \frac{1}{4})^{-1} \quad \text{for} \quad \frac{1}{4} < \alpha^2 \leq 17/8.$$



Now  $\alpha^4 - 2\alpha^2 + 3 = (\alpha^2 + 1)^2 + 2 > 0$  and so  $\psi$  decreases as  $k$  increases; hence

$\psi \leq 1 - (\alpha^4 - 2\alpha^2 + 3)(4 - \alpha^2)^{-2} = \psi_1$ , say, for  $17/8 \leq \alpha^2 < 4$  and

$\psi \leq 1 - (\alpha^4 - 2\alpha^2 + 3)(\alpha^2 - \frac{1}{4})^{-2} = \psi_2$ , say, for  $\frac{1}{4} < \alpha^2 \leq 17/8$ .

Now  $\partial\psi_1/\partial\alpha = 0$  for  $\alpha = 0, \pm 3^{-1/2}$  and so, for  $17/8 \leq \alpha^2 < 4$ ,  $\psi_1$  is a decreasing function of  $\alpha$  and  $\psi \leq \psi_1(\sqrt{(17/8)}) = 16/225$ . Similarly  $\psi_2$  is an increasing function of  $\alpha$  for  $\frac{1}{4} < \alpha^2 \leq 17/8$  and  $\psi \leq \psi_2(\sqrt{(17/8)}) = 16/225$ .

When  $\alpha^2 = 17/8$  and  $k = 64/225$  then

$$g(x) = 1 - (8x^2 - 17)^2/225$$

and

$$g(\pm 2) = 0, \quad g(\pm \frac{1}{2}) = 0.$$

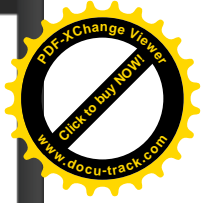
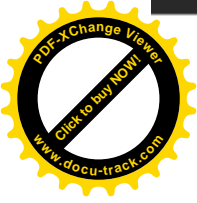
If  $\alpha_4 = 0$  then we take  $g(x) = 1 - k(x - \alpha)^2$  and we shall have either  $g(\frac{1}{2}) = 0$  and  $g(2) \leq 0$  or  $g(2) = 0$  and  $g(\frac{1}{2}) \leq 0$ . In the first case

$$g(x) = 1 - (x - \alpha)^2(\alpha - \frac{1}{2})^{-2} \quad \text{where } \alpha \leq 5/4$$

and  $\psi = 1 - (1 + \alpha^2)(\alpha - \frac{1}{2})^{-2}$ ; this makes  $\psi$  negative. In the second case  $g(x) = 1 - (x - \alpha)^2(2 - \alpha)^{-2}$  and  $5/4 \leq \alpha$ ;  $\psi = (3 - 4\alpha)(2 - \alpha)^{-2}$  and is greatest for  $\alpha = 5/4$ , being then also negative. Hence  $\sup \psi$  is not less than  $16/225$  and so  $L \geq 16/225$ .

However, we can construct a discrete distribution with probabilities  $p$  at  $\sqrt{(17/8)}$ ,  $((16/225) - p)$  at  $-\sqrt{(17/8)}$ , and  $q, r$  at two of the points  $\pm 2, \pm \frac{1}{2}$ , to give the required moments and also give  $P(T) = 16/225$ . (For example,  $p = (8\sqrt{17} + 6\sqrt{8})/225 \sqrt{17}$ ,  $q = 37/225$  at  $2$ ,  $r = 172/225$  at  $-\frac{1}{2}$ .) Hence  $L \leq 16/225$  and so  $L = 16/225$ .

Since  $g(x)$  is equal to  $\chi_T(x)$  for more than four values of  $x$  when  $\psi$  attains its maximum, we have that the support hyperplane  $a_0 + a_2 + 3a_4 = L$  is linearly dependent on more than four support hyperplanes and so is linearly dependent on support hyperplanes in an infinity of ways. Hence there is an infinity of distributions which give the value  $L$  for  $P(T)$ , and among these there is just one symmetrical one (with probabilities  $8/225$  at  $\pm\sqrt{(17/8)}$ ,  $86/225$  at



$\pm \frac{1}{2}$ ,  $37/450$  at  $\pm 2$ ). We could, in fact, have simplified the above working by using the symmetry of the problem since if  $f_1(x)$  gives  $P(T) = p_1$  then  $f_1(-x)$  and the symmetric frequency function  $\frac{1}{2}(f_1(x) + f_1(-x))$  both also give  $P(T) = p_1$  and satisfy the same moment conditions.

Since the polynomial  $g(x)$  giving  $L$  is equal to  $\chi_T(x)$  for  $x = \pm \frac{1}{2}$ ,  $\pm 2$  as well as at  $\alpha$  and  $\beta$ , it follows that we should have arrived at the same  $g(x)$  and so at  $L$  by taking  $\beta$  as one of the values  $\pm \frac{1}{2}$ ,  $\pm 2$  instead of the value at which  $g(x)$  attained maximum value of unity; it may be verified, however, that the analysis is more complicated in these cases.

## 2.5 Single interval; first and second moments given

To illustrate the technique used above when values are given algebraically and not numerically, we consider the case when  $\mu'_1 = 0$ ,  $\mu'_2 = 1$ , and  $T$  is the interval  $-\alpha \leq x \leq \beta$  with  $0 < \alpha \leq \beta$ ; i.e.  $T$  is an interval containing the mean and extending from the mean at least as far to the right as to the left.

We can satisfy the moment conditions with probability unity at 0 and zero at infinity, and this gives  $U = 1$ . To find  $L$  we take  $T$  now as  $-\alpha < x < \beta$  and consider quadratic polynomials  $g(x) = a_0 + a_1x + a_2x^2$  satisfying

$$g(x) \leq \chi_T(x) = 0 \quad (x \leq -\alpha \text{ or } \beta \leq x)$$

and

$$g(x) \leq \chi_T(x) = 1 \quad (-\alpha < x < \beta).$$

By increasing first  $a_0$  until  $g(x) = \chi_T(x)$  at  $x = \gamma$  and then adding a positive multiple of  $(x - \gamma)^2$  we may suppose that  $g(x) = \chi_T(x)$  for  $x = \gamma$ ,  $x = \delta$ . If  $L > 0$  then we may take  $-\alpha < \gamma < \beta$  (by interchanging  $\gamma$  and  $\delta$  if necessary); since  $g(x)$  is quadratic and therefore  $g(x) = 0$  has not more than two real roots, we must have  $\delta = -\alpha$  or  $\delta = \beta$ .

If  $\delta = -\alpha$  then  $g(x) = 1 - (x - \gamma)^2(\alpha + \gamma)^{-2}$  and  $g(\beta) = 1 - (\beta - \gamma)^2(\alpha + \gamma)^{-2} \leq 0$  so that  $\beta - \gamma \geq \gamma + \alpha$ . We have  $\psi = 1 - (1 + \gamma^2)(\alpha + \gamma)^{-2}$  and  $d\psi/d\gamma = 2(1 - \alpha\gamma)(\alpha + \gamma)^{-3}$ . Hence  $\psi$  increases for  $-\alpha < \gamma < \alpha^{-1}$  and if  $2 \leq \alpha(\beta - \alpha)$  the

maximum value of  $\psi$  is  $\alpha^2/(\alpha^2 + 1)$  for  $\gamma = \alpha^{-1}$ . If, however,  $2 > \alpha(\beta - \alpha)$  then  $\psi$  has its greatest value for  $\gamma = \frac{1}{2}(\beta - \alpha)$ , and the value is  $4(\alpha\beta - 1)/(\alpha + \beta)^2$ .

If  $\delta = +\beta$  then  $g(x) = 1 - (x - \gamma)^2/(\beta - \gamma)^2$ , and since  $g(-\alpha) \leq 0$  we have  $\alpha + \gamma \geq \beta - \gamma$ .

$\psi = 1 - (1 + \gamma^2)/(\beta - \gamma)^2$  and  $d\psi/d\gamma = -2(1 + \gamma\beta)/(\beta - \gamma)^3$ . Since  $2\gamma \geq \beta - \alpha \geq 0$ ,  $\psi$  decreases as  $\gamma$  increases, and the greatest value of  $\psi$  is for  $\gamma = \frac{1}{2}(\beta - \alpha)$ , being then  $4(\alpha\beta - 1)/(\alpha + \beta)^2$ .

Now

$$\frac{\alpha^2}{\alpha^2 + 1} - \frac{4(\alpha\beta - 1)}{(\alpha + \beta)^2} = \frac{(\alpha\beta - \alpha^2 - 2)^2}{(\alpha^2 + 1)(\alpha + \beta)^2},$$

and so when  $\alpha^2/(\alpha^2 + 1)$  is a possible value for  $\psi$ , i.e.

$$\alpha(\beta - \alpha) \geq 2, \text{ then } L = \alpha^2/(\alpha^2 + 1);$$

$$\text{when } \alpha(\beta - \alpha) \leq 2, \text{ then } L = 4(\alpha\beta - 1)/(\alpha + \beta)^2.$$

We assumed at the start that  $L$  was greater than 0; if  $\alpha\beta \leq 1$  then  $\alpha(\beta - \alpha) \leq 2$  and the above argument gives  $L \leq 0$ ; this is a contradiction and so, for  $\alpha\beta \leq 1$ , we have  $L = 0$ .

The above result was first given (with a general value for  $\mu_2$ ) by Selberg (1940) who used a special method depending on Schwarz's inequality.

If we take  $\alpha = \beta$  then we have  $L = 1 - \alpha^{-2}$ ; this means that for the complement of  $T$ , i.e. the set  $|x| \geq \alpha$ , we have  $U = \alpha^{-2}$ . This result is one of the oldest in the subject and was discovered by Bienaymé in 1853. Tchebychef rediscovered it in 1867, and the prestige of the greater mathematician has resulted in his name being generally applied to it and to the whole subject which has developed from it. (The name of "Tchebychef's inequalities" is also given to the inequalities for the case when  $T$  is  $x \leq k$ . These were stated without proof by Tchebychef in 1874 but were proved ten years later by Markov and Stieltjes independently. We shall consider these in the next section.)

## 2.6 Half line; ordinary moments given

In this section we take for  $T$  the set  $0 \leq x$  and assume a knowledge of the moments  $\mu'_1, \dots, \mu'_{2n}$ . (In previous examples we

have chosen the origin and scale so that  $\mu'_1$  was 0 and  $\mu'_2$  was unity; in the present example it leads to a more symmetrical presentation if instead we make the origin the end-point of  $T$ . For applications with given numerical values, it may be better to revert to the previous usage.)

It will suffice to determine  $U$  since, as we shall see later, the procedure simultaneously determines  $L$ .

We consider polynomials  $g(x) = a_0 + a_1 x + \dots + a_{2n} x^{2n}$ , and by adding successively negative multiples of expressions such as  $1, (x - \alpha)^2, (x - \alpha)^2 (x - \beta)^2, \dots$ , we can suppose that  $g(x) = \chi_T(x)$  for at least  $n + 1$  values of  $x$ , while  $g(x) \geq \chi_T(x)$  generally. Since there must be a maximum between each pair of minima of  $g(x)$ , and since  $g(x)$  cannot have more than  $2n - 1$  turning-points, it follows that  $x = 0$  must be one of the points at which  $g(x) = \chi_T(x)$ . We now show that there exist a discrete distribution having probability at 0 and  $r$  other points  $x_1, \dots, x_r$  ( $r \leq n$ ), and a polynomial of the required kind such that  $g(x) = \chi_T(x)$  only for  $x = 0, x_1, \dots, x_r$ .

The matrix

$$M = \begin{pmatrix} 1 & \mu'_1 & \dots & \mu'_n \\ \cdot & \cdot & \dots & \cdot \\ \mu'_n & \mu'_{n+1} & \dots & \mu'_{2n} \end{pmatrix}$$

is, by hypothesis, that of a positive definite form; let  $p$  be such that

$$M_p = \begin{pmatrix} 1-p & \mu'_1 & \dots & \mu'_n \\ & & \dots & \\ \mu'_n & \mu'_{n+1} & \dots & \mu'_{2n} \end{pmatrix}$$

has determinant zero. Since

$$\begin{vmatrix} 1-p & \dots & \mu'_{n-1} \\ & \dots & \\ \mu'_{n-1} & \dots & \mu'_{2n-2} \end{vmatrix} \begin{vmatrix} \mu'_2 & \dots & \mu'_{n+1} \\ & \dots & \\ \mu'_{n+1} & \dots & \mu'_{2n} \end{vmatrix} - \begin{vmatrix} \mu'_1 & \dots & \mu'_n \\ & \dots & \\ \mu'_n & \dots & \mu'_{2n-1} \end{vmatrix}^2 = 0$$

(by the Jacobi Ratio Theorem: see, e.g. Mirsky (1955), p. 25), we have

$$\begin{vmatrix} 1-p & \dots & \mu'_{n-1} \\ & \dots & \\ \mu'_{n-1} & \dots & \mu'_{2n-2} \end{vmatrix} \geq 0.$$

Similarly every principal minor of  $M_p$  has non-negative determinant, so that  $p \leq 1$ , and also  $M_p$  is the matrix of a positive semi-definite form. So too is the matrix

$$\frac{M_p}{(1-p)} = \begin{pmatrix} 1 & \mu'_1/(1-p) & \dots & \mu'_n/(1-p) \\ \mu'_n/(1-p) & \mu'_{n+1}/(1-p) & \dots & \mu'_{2n}/(1-p) \end{pmatrix}.$$

Now, as stated without proof in Section 2.2, there exists a distribution with moments  $\mu'_1/(1-p), \dots, \mu'_{2n}/(1-p)$ . Since the matrix  $M_p/(1-p)$  is that of a semi-definite form, there exist also real numbers  $a_0, \dots, a_n$  such that, for this distribution,  $E((a_0 + \dots + a_n x^n)^2) = 0$ . Hence the distribution is discrete with probabilities  $q_1, \dots, q_r$  at  $x_1, \dots, x_r$  which are real zeros of  $a_0 + \dots + a_n x^n$ . We do not need to discuss the relation between  $r$  and  $n$ , but we may note in passing that if  $r < n$  then  $\mu'_1, \dots, \mu'_{2n}$  are moments of a discrete distribution with at most  $n$  values. This is possible even if the determinant of  $M$  is positive, since we may change  $\mu'_{2n}$  by adding zero probability at infinity to the discrete distribution which we have constructed.

Hence there is a discrete distribution with probabilities  $p$  at 0,  $(1-p)q_i$  at  $x_i$  ( $i=1, \dots, r$ ), with moments  $\mu'_1, \dots, \mu'_{2n}$ .

Suppose that  $x_1, \dots, x_s$  are negative, and  $x_{s+1}, \dots, x_r$  are positive ( $0 \leq s \leq r$ , with  $s=0$  or  $s=r$  meaning that all the  $x_i$  have the same sign). Then a polynomial  $g(x)$  with values 0 at  $x_1, \dots, x_s$ , values 1 at 0,  $x_{s+1}, \dots, x_r$ , and satisfying  $g(x) \geq 0$  ( $x < 0$ ),  $g(x) \geq 1$  ( $0 \leq x$ ), is

$$1 + x(x - x_{s+1})^2 \dots (x - x_r)^2 Q(x),$$

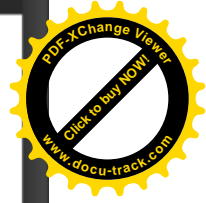
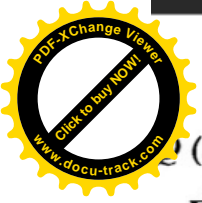
where  $Q(x)$  is a polynomial to be determined. (This is for  $s > 0$ ; if  $s=0$  we may take simply  $g(x) = 1$ .) We need

$$1 + x_i(x_i - x_{s+1})^2 \dots (x_i - x_r)^2 Q(x_i) = 0$$

and

$$\frac{1}{x_i} + \frac{2}{x_i - x_{s+1}} + \dots + \frac{2}{x_i - x_r} + \frac{Q'(x_i)}{Q(x_i)} = 0 \quad (i=1, \dots, s).$$

These equations determine  $Q(x_i)$  and  $Q'(x_i)$  and then



$$Q(x) =$$

$$\sum_{i=1}^s Q(x_i) P_i(x) + \sum_{i=1}^s \{Q'(x_i) - Q(x_i) P_i'(x_i)\} P_i(x) (x - x_i),$$

$$\text{where } P_i(x) = \frac{(x - x_1)^2 \dots (x - x_{i-1})^2 (x - x_{i+1})^2 \dots (x - x_s)^2}{(x_i - x_1)^2 \dots (x_i - x_{i-1})^2 (x_i - x_{i+1})^2 \dots (x_i - x_s)^2},$$

with obvious modifications in the cases  $i=1$ ,  $i=s$ .

The distribution which we have constructed now gives

$$U = p + q_{s+1} + \dots + q_r.$$

The polynomial  $1 - g(x)$  satisfies the requirements on  $g(x)$  which are needed in determining  $U$  when  $T$  is  $x \leq 0$ . Hence, since the distribution depends only on  $0, x_1, \dots, x_r, U$ , for this problem, is  $q_1 + \dots + q_s + p$ . But  $U$  for this problem is  $1 - L$  for the original problem and so

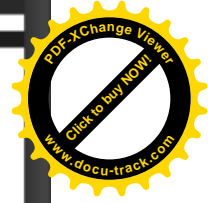
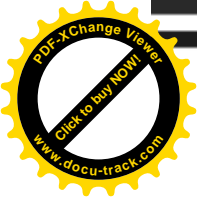
$$L = 1 - (q_1 + \dots + q_s + p) = q_{s+1} + \dots + q_r.$$

For example, with  $\mu'_1 = -1$ ,  $\mu'_2 = 2$ ,  $\mu'_3 = -4$ ,  $\mu'_4 = 10$  (equivalent to  $\mu'_1 = \mu'_3 = 0$ ,  $\mu'_2 = 1$ ,  $\mu'_4 = 3$  and  $T$  as  $1 \leq x$ ), then  $p = \frac{1}{2}$  and  $-2, 4, -8, 20$  are the first four moments of the distribution with probability 1 at  $-2$  and zero at infinity. Hence  $q_1 = \frac{1}{2}$ ,  $L = 0$  and  $U = \frac{1}{2}$ . In this case  $g(x) = \frac{1}{4}(x+2)^2$  and is of degree less than  $2n (= 4)$ .

If  $\mu'_1 = 0$ ,  $\mu'_2 = \frac{1}{2}$ ,  $\mu'_3 = 0$ ,  $\mu'_4 = \frac{1}{2}$  then  $p = \frac{1}{2}$ , and for the distribution with moments  $0, 1, 0, 1$  we have  $E((x^2 - 1)^2) = 0$ , so that the distribution has probabilities  $\frac{1}{2}$  at  $x = -1$  and  $x = 1$ . This gives  $L = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ ,  $U = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$ . In finding  $U$  we have  $g(x) = \frac{1}{4}(2x^4 - x^3 - 4x^2 + 3x + 4)$ , so that the degree of  $g(x)$  is  $2n (= 4)$ .

Taking  $T$  as  $-2 \leq x$  and the moments those of the normal distribution with zero mean and unit variance, we have the following results —

$n$	$L$	$U$
1	.8	1
2	.8947 ...	1
3	.9016 ...	1
4	.9106 ...	.9998 ...



If  $T$  is  $-1 \leq x$  then the same moments give —

$n$	$L$	$U$
1	.5	1
2	.5	1
3	.6038 ...	.9788 ...
4	.6209 ...	.9739 ...

It can be seen that the extra work involved in using higher moments adds little to the information yielded.

In the above work it has been assumed that a set of consecutive moments is given; we could in fact deal in a similar way with any set of moments, except that the criteria for a distribution to consist of discrete probabilities are less easy to formulate.

The fact that the values for  $L$  and  $U$  are best possible has been proved by Marshall and Olkin (1961), using the ideas of the theory of games, for the case discussed in this section as well as for some other cases.

## 2.7 Absolute moments given

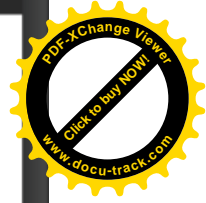
If we are given absolute moments of a distribution about a value  $a$  (which we shall assume to be zero) then we may regard the distribution as symmetrical about  $O$ ; and  $T$ , which will be defined initially only in terms of non-negative  $x$ , is the symmetrical set of intervals obtained by reflection in the point  $x = 0$ . If the moments given are those of orders  $1, \dots, n$  then we take

$$g(x) = a_0 + a_1|x| + \dots + a_n|x|^n. \quad (2.7.1)$$

Because of the symmetry of  $\chi_T(x)$  it is sufficient to have  $g(x) \geq \chi_T(x)$  or  $g(x) \leq \chi_T(x)$  (according as to whether it is  $U$  or  $L$  which we are finding) only for non-negative  $x$ ; we can then drop the modulus signs in (2.7.1), and the difficulties which might be caused by the use of moduli are obviated.

We shall consider only the case when  $T$  is  $k \leq x$  (corresponding to the case  $|x| \geq k$  in general) and we then have, in finding  $U$ , to choose  $g(x)$  so that

$$g(x) \geq 0 \quad \text{for } 0 \leq x < k, \quad g(x) \geq 1 \quad \text{for } k \leq x.$$



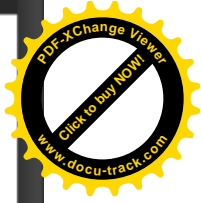
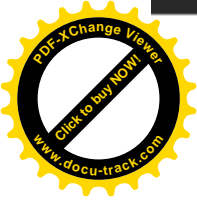
Since we are concerned only with positive values of  $x$  we can subtract positive multiples of terms such as  $(x - \alpha_1)^2 \dots (x - \alpha_r)^2$  or  $x(x - \alpha_1)^2 \dots (x - \alpha_r)^2$ ; if  $n$  is even, say  $n = 2m$ , this means that  $g(x) = \chi_T(x)$  for at least  $m + 1$  positive values of  $x$  or for 0 and at least  $m$  other values. Since  $g(x)$  tends to infinity as  $x$  tends to infinity (except in the trivial case when  $g(x) = 1$ ), then if  $g(x)$  has minima at the positive values there are in each case at least  $2m + 1$  turning-points, and this is impossible; hence we must have  $g(k) = \chi_T(k)$ . This holds also for the case when  $n$  is odd. By the argument used in Section 2.6 we can obtain a discrete distribution which has non-zero probability only at some or all of the points at which  $g(x) = \chi_T(x)$  and possibly zero probability at infinity, and the values of  $L$  and  $U$  are obtained from this distribution.

If we are given not a consecutive set of moments but those of orders  $i_1, \dots, i_n$  then Wald (1939) has shown that the same results obtain, but a more refined argument about the changes of sign of polynomials containing only certain powers of  $x$  has to be used, and it is less easy to show the existence of the discrete distribution. Isii (1959b) has dealt with the problem on the lines which have been followed here.

As an example, suppose  $k = 2$  and the first two, three, or four absolute moments of the normal distribution are given.

For  $n = 2$  we take probabilities  $p$  at  $x = 2$  and  $q$  at  $x = \alpha$  to satisfy  $1 = p + q$ ,  $\sqrt{2/\pi} = 2p + q\alpha$ ,  $1 = 4p + q\alpha^2$  and so we have  $\sqrt{2/\pi} - 2 = q(\alpha - 2)$ ,  $1 - 2\sqrt{2/\pi} = q\alpha(\alpha - 2)$ , whence  $\alpha = .4955 \dots$ ,  $p = .2010 \dots$ ,  $q = .7990 \dots$ , and finally  $L = .7990 \dots$ ,  $U = 1$ .

For  $n = 3$ , if we take probabilities  $p$  at  $x = 2$  and  $q$  at  $x = \alpha$  as above, then  $8p + q\alpha^3 = 1.7051 \dots > 2\sqrt{2/\pi} = \nu_3$ , and this inequality is only made worse if we try to add zero probability at infinity. Hence we must try instead  $p$  at  $x = 2$ ,  $q$  at  $x = 0$  and  $r$  at  $x = \alpha$ , which give  $p = .1735 \dots$ ,  $q = .1620 \dots$ ,  $r = .6645 \dots$  and  $\alpha = .6785 \dots$  so that finally  $L = .8265 \dots$ ,  $U = 1$ . Since these values give  $16p + r\alpha^4 = 2.916 \dots < 3 = \nu_4$ , we merely add probability zero at infinity to give the correct fourth moment when  $n = 4$  and so obtain the same values of  $L$  and  $U$ . This means that in  $g(x)$  we take  $a_4$  as zero and so continue to use a cubic polynomial.



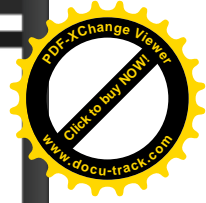
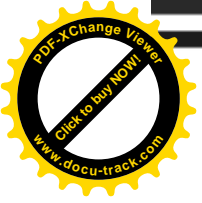
The corresponding values of  $L$  and  $U$  for ordinary moments with intervals  $2 \leq |x|$  are

$n$	$L$	$U$
2	.75	1
3	.75	1
4	.8182 ...	1

It will be observed that the absolute moments add little information when  $n = 4$ , but for  $n = 2$  or for  $n = 3$ , when the third ordinary moment adds no information, a knowledge of the first and third absolute moments gives a considerable improvement.

## 2.8 Finite interval

We have so far considered variables whose distribution can extend to infinity in each direction, and where we have the possibility of varying the moment of highest order which we are given by adding zero probability at infinity. If the distribution is restricted to lie in a finite interval then the methods given in Sections 2.6 and 2.7 still apply, except that now we may have non-zero probability at the ends of the interval. Suppose we are given the first  $n$  moments  $\mu'_1, \dots, \mu'_n$ ,  $f(x)$  is zero outside the interval  $b \leq x \leq c$ , and  $T$  is  $k \leq x \leq c$ . The inequality  $g(x) \geq \chi_T(x)$  (if we are finding  $U$ ) has now to be satisfied only for  $b \leq x \leq c$ ; hence we may reduce  $E(g(x))$  by adding negative multiples of terms such as  $(x-b)(x-\alpha_1)^2 \dots$  or  $(c-x)(x-\alpha_1)^2 \dots$  or  $(x-b)(c-x)(x-\alpha_1)^2 \dots$ . Note that the coefficient of  $x$  in  $g(x)$  may be of either sign, in contrast to the case when the interval is infinite. By an argument similar to that in Section 2.6 we shall have  $g(k) = \chi_T(k)$ . If  $n$  is odd ( $n = 2m - 1$ ) then if we have  $g(b) = \chi_T(b)$  or  $g(c) = \chi_T(c)$  we can have also  $m - 1$  other values at which  $g(x) = \chi_T(x)$ ; if  $g(b) \neq \chi_T(b)$  and  $g(c) \neq \chi_T(c)$  then we still cannot be sure of more than  $m - 1$  other values, but in this case there will have to be some special relation between the moments for the  $2m$  moment conditions (including the moment of order zero) to be satisfied by  $m - 1$  values of  $x$  and  $m$  probabilities. Similarly if  $n$  is even ( $n = 2m$ ) we may have  $g(x) = \chi_T(x)$  either at  $k$  and  $m$  other values different from  $b$  and  $c$ , or at  $b, k, c$  and  $m - 1$  other values.



If  $b = -1$ ,  $c = 1$ ,  $k = 0$ ,  $\mu'_1 = 0$ ,  $\mu'_2 = \frac{1}{3}$  (moments of the rectangular distribution) then we construct distributions with probabilities  $p$  at  $x = -1$  and  $q$  at  $x = 0 (= k)$  or else  $p$  at  $x = -1$ ,  $q$  at  $x = 0$  and  $r$  at  $x = 1$ . The first possibility leads to no solution, while the second gives  $p = r = \frac{1}{6}$ ,  $q = \frac{2}{3}$ . Hence  $U = \frac{5}{6}$  and, by symmetry,  $L = \frac{1}{6}$ .

Use of further moments gives  $L$ ,  $U$  as follows —

$n$	$L$	$U$
3	.1667 ...	.8333 ...
4	.2778 ...	.7222 ...
5	.2778 ...	.7222 ...

If  $T$  is  $\frac{1}{2} \leq x \leq 1$  we obtain

$n$	$L$	$U$
2	.4286 ...	1
3	.5152 ...	1
4	.5335 ...	.9286 ...
5	.5472 ...	.9247 ...

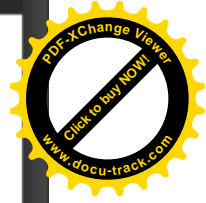
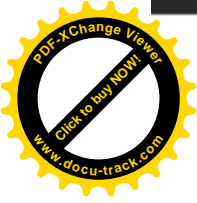
As in other examples, higher moments yield a rapidly diminishing return on the labour involved in using them.

## 2.9 Expectations of general functions given

When we are given not moments of a distribution but the expected values of functions other than powers of  $x$ , then the results we obtain will depend to a great extent on what these functions are. We can no longer, as in the sections above, use the familiar properties of polynomials.

If  $E(\phi(x))$  is given and we require  $U$  when  $T$  is  $k \leq x$ , then we consider  $g(x) = a_0 + a_1 \phi(x)$  and require  $a_0 + a_1 \phi(x) \geq \chi_T(x)$ . This means that unless  $a_1 = 0$ , leading to the trivial solution  $U = 1$ ,  $\phi(x)$  must have a finite lower bound. We consider some special functions satisfying this condition.

If  $\phi(x) = e^x$  we note that  $E(e^x)$  can be increased by the addition of zero probability at  $+\infty$ , while a finite probability at a large negative value of  $x$  will only affect  $E(e^x)$  by a small amount, so that in this case we need also the limiting concept of finite probability at  $-\infty$  which leaves  $E(e^x)$  unaltered.



If  $E(e^x) \leq e^k$  we can take probabilities  $E(e^x)/e^k$  at  $x = k$  and  $1 - (E(e^x)/e^k)$  at  $-\infty$  to give  $U = E(e^x)/e^k$ ; if  $E(e^x) \geq e^k$  then  $U = 1$ . In this simple case the reader will find it instructive to sketch the convex set and support line in the  $(a_0, a_1)$  plane.

For the normal distribution with zero mean and unit variance we have  $E(e^x) = e^{x^2/2}$  and for  $k > \frac{1}{2}$  we have  $U = e^{k^2/2}$ . From the moments  $\mu'_1 = 0$ ,  $\mu'_2 = 1$  we have (see Section 2.6)  $U = 1/(k^2 + 1)$  for  $k > 0$ , so that knowledge of  $E(e^x)$  (which implies much greater knowledge of the distribution for large positive  $x$ ) gives more information when  $k$  is large and positive, but less for  $k < 2.44 \dots$ . If  $\phi(x) = e^{x^2/4}$  then we can no longer have finite probability at  $-\infty$ , but if  $E(e^{x^2/4}) < e^{k^2/4}$  we can have probabilities

$$\begin{aligned} & (e^{k^2/4} - E(e^{x^2/4})) / (e^{k^2/4} - 1) \quad \text{at } x = 0 \text{ and} \\ & (E(e^{x^2/4}) - 1) / (e^{k^2/4} - 1) \quad \text{at } x = k, \text{ giving} \\ & U = (E(e^{x^2/4}) - 1) / (e^{k^2/4} - 1). \end{aligned}$$

Again a sketch of the  $(a_0, a_1)$  plane will be helpful.

For the normal distribution as above, we have  $E(e^{x^2/4}) = \sqrt{2}$ , giving  $U = (\sqrt{2} - 1) / (e^{k^2/4} - 1)$ . This is an improvement on  $U = 1/(k^2 + 1)$  for  $k > 2.23 \dots$

For the case when the expectations of two functions are given we shall consider particular conditions which lead to a result of von Mises (1939). We suppose that  $f(x)$  is non-zero only for  $0 \leq x \leq d$ , that  $\phi_1(x)$  and  $\phi_2(x)$  are the functions whose expectations are given and are such that

$$\begin{aligned} \phi_1(0) = \phi_2(0) = 0, \quad \phi'_1(x) > 0, \quad \phi'_2(x) > 0, \text{ and} \\ \phi''_2(x) \phi'_1(x) - \phi''_1(x) \phi'_2(x) > 0. \end{aligned}$$

The conditions  $\phi'_1(x) > 0$ ,  $\phi'_2(x) > 0$  ensure that there is just one value of  $x$  for a given value of  $\phi_1(x)$  or  $\phi_2(x)$ , while the condition

$$\phi''_2(x) \phi'_1(x) - \phi''_1(x) \phi'_2(x) > 0$$

means that  $\phi'_2(x)/\phi'_1(x)$  is strictly increasing, so that, for given  $a_1, a_2$  not both zero, the equation  $a_1 \phi'_1(x) + a_2 \phi'_2(x) = 0$  has at most one solution in  $x$ . Hence  $g(x) = a_0 + a_1 \phi_1(x) + a_2 \phi_2(x)$  takes a given

value for at most two values of  $x$ , for each set of values of  $a_0, a_1, a_2$  ( $a_1, a_2$  not both zero). We define  $d_1, d_2$  by the equations

$$\frac{\phi_2(d_1)}{\phi_1(d_1)} = \frac{E(\phi_2(x))}{E(\phi_1(x))}$$

and

$$\begin{vmatrix} \phi_1(d_2) & \phi_2(d_2) & 1 \\ E(\phi_1(x)) & E(\phi_2(x)) & 1 \\ \phi_1(d) & \phi_2(d) & 1 \end{vmatrix} = 0.$$

In geometrical terms, the curve  $u = \phi_1(x)$ ,  $v = \phi_2(x)$  in co-ordinates  $(u, v)$  between the origin  $O$  and the point  $D(\phi_1(d), \phi_2(d))$  and the chord  $OD$  together enclose a convex region, and the point  $G(E(\phi_1(x)), E(\phi_2(x)))$  lies inside or on the boundary of this region (this is most easily seen by approximating to the p.d.f. used to give the expected values by a discrete distribution). Hence the points  $D_1(\phi_1(d_1), \phi_2(d_1))$  and  $D_2(\phi_1(d_2), \phi_2(d_2))$ , which lie on  $OG, DG$  respectively, lie in the arc  $OD$ .

If  $0 \leq k \leq d_2$ , let  $k'$  be defined by

$$\begin{vmatrix} \phi_1(k') & \phi_2(k') & 1 \\ E(\phi_1(x)) & E(\phi_2(x)) & 1 \\ \phi_1(k) & \phi_2(k) & 1 \end{vmatrix} = 0,$$

i.e. the points  $K(\phi_1(k), \phi_2(k))$ ,  $K'(\phi_1(k'), \phi_2(k'))$  and  $G$  are collinear. We obtain the correct expectations  $E(\phi_1(x))$ ,  $E(\phi_2(x))$  with probabilities

$$\begin{aligned} \frac{-E(\phi_1(x)) + \phi_1(k')}{-\phi_1(k) + \phi_1(k')} &= \frac{\phi_2(k') - E(\phi_2(x))}{\phi_2(k') - \phi_2(k)} \quad \text{at } x = k \text{ and} \\ \frac{E(\phi_1(x)) - \phi_1(k)}{\phi_1(k') - \phi_1(k)} &= \frac{E(\phi_2(x)) - \phi_2(k)}{\phi_2(k') - \phi_2(k)} \quad \text{at } x = k'. \end{aligned}$$

Now if  $a_0, a_1, a_2$  are such that  $g(k) = 1$ ,  $g(k') = 0$ ,  $g'(k') = 0$ , we have  $g(x) > 0$  for  $x \neq k'$  and also  $g(0) = a_0 =$

$$\frac{\phi_1(k')\phi_2'(k') - \phi_2(k')\phi_1'(k')}{\phi_1(k')\phi_2'(k') - \phi_2(k')\phi_1'(k') + \phi_2(k)\phi_1'(k') - \phi_1(k)\phi_2'(k')}.$$

The function  $\phi(x) = \phi_2(x)\phi_1'(k') - \phi_1(x)\phi_2'(k')$  has zero derivative at  $x = k'$ , and this is its only stationary point. It is 0 for  $x = 0$  and

consequently  $a_0 = \phi(k')/(\phi(k') - \phi(k)) > 1$ , so that  $g(x) \geq 1$  for  $0 \leq x \leq k$ .

Hence  $U = (\phi_1(k') - E(\phi_1(x)))/(\phi_1(k') - \phi_1(k))$  and, from the construction,  $L = 0$ . A similar argument holds for  $d_1 < k < d$ , giving

$$L = (E(\phi_1(x)) - \phi_1(k))/(\phi_1(k') - \phi_1(k)), \quad U = 1.$$

For  $d_2 < k < d$ , we take probabilities

$$\begin{aligned} p_1 &= \frac{(E(\phi_1(x)) - \phi_1(d))(E(\phi_2(x)) - \phi_2(k)) - (E(\phi_1(x)) - \phi_1(k))(E(\phi_2(x)) - \phi_2(d))}{\phi_1(d)\phi_2(k) - \phi_1(k)\phi_2(d)} & \text{at } x = 0, \\ p_2 &= \frac{E(\phi_2(x))\phi_1(d) - E(\phi_1(x))\phi_2(d)}{\phi_1(d)\phi_2(k) - \phi_1(k)\phi_2(d)} & \text{at } x = k \\ p_3 &= \frac{E(\phi_1(x))\phi_2(k) - E(\phi_2(x))\phi_1(k)}{\phi_1(d)\phi_2(k) - \phi_1(k)\phi_2(d)} & \text{at } x = d \end{aligned}$$

and choose  $a_0, a_1, a_2$  so that  $1 = g(0) = g(k)$ ,  $0 = g(d)$ ; since  $g(x)$  has a single turning-point this gives  $g(x) \geq \chi_T(x)$ , and we have  $L = p_1$ ,  $U = p_1 + p_2$ .

If the distribution has infinite range we can let  $d$  tend to infinity in the above values. If  $\phi_1(d)$  and  $\phi_2(d)$  remain finite as  $d$  tends to infinity, we may need positive probability at infinity; if  $\phi_2(d)/\phi_1(d)$  tends to infinity as  $d$  tends to infinity then  $E(\phi_2(x))$  may be affected by zero probability at infinity.

As an example, if  $\phi_1(x) = e^{x^{1/8}} - 1$ ,  $\phi_2(x) = e^{x^{1/4}} - 1$ , and  $E(\phi_1(x)) = \sqrt{(\frac{4}{3})} - 1$ ,  $E(\phi_2(x)) = \sqrt{2} - 1$  (the values for the normal distribution with zero mean and unit variance), then  $d_1 = \sqrt{(8 \log((\sqrt{6} - 2)/(2 - \sqrt{3})))} = 2.034 \dots$ ,  $d_2 = \sqrt{(8 \log(2(\sqrt{3})))} = 1.073 \dots$ . For  $k = 3$  we have  $k' = .924 \dots$  which gives  $L = .979 \dots$

Using a single function  $\phi(x)$ , we have  $L = 1 - E(\phi(x))/\phi(k)$  which gives  $L = .926 \dots$  for  $\phi_1(x)$  and  $L = .951 \dots$  for  $\phi_2(x)$ ,  $U$  being 1 in every case.

## 2.10 Mean range given

The information about the distribution of the variate  $x$  in the previous sections of this chapter has all been in the form of the expectations of various functions of  $x$ , these functions being independent of the particular distribution. In this section we take as

datum the mean range of a sample of given size; this can be represented only as the expectation of a function of the distribution function, and the method used hitherto no longer applies. Instead we use a method which consists essentially of replacing the given distribution by one more closely grouped about the median. It will be seen that the inequality obtained is different in type from those in the earlier sections.

If  $w$  is the mean range in samples of  $n$  from the population with distribution function  $F(x)$ , then

$$\begin{aligned} w &= \int_{-\infty}^{\infty} \{1 - F^n - (1 - F)^n\} dx \\ &= \int_{-\infty}^{\infty} R(F) dx, \text{ say.} \end{aligned}$$

(For a proof, see, for example, Kendall and Stuart (1958, 1963), p. 339.)

Let  $W(m) = \sum_{i=1}^{m-1} R\left(\frac{i}{m}\right)$ ; since  $\frac{d^2 R}{dF^2} < 0$  for  $0 < F < 1$ , we have,

by comparing areas under chords and tangents to the graph of  $R(F)$  and under the graph itself,

$$m \int_{1/(2m)}^{1-1/(2m)} R(u) du < W(m) < m \int_0^1 R(u) du.$$

Hence

$$\begin{aligned} W(m+1) - W(m) &> (m+1) \int_{1/\{2(m+1)\}}^{1-1/\{2(m+1)\}} R(u) du - m \int_0^1 R(u) du \\ &= \frac{2}{n+1} \left\{ m - (m+1) \left( \frac{2m+1}{2m+2} \right)^{n+1} + (m+1) \left( \frac{1}{2m+2} \right)^{n+1} \right\} \\ &> \frac{2}{n+1} \left\{ m - (m+1) \left( \frac{2m+1}{2m+2} \right)^3 \right\} > 0; \end{aligned}$$

also  $W(1) = 0$ , and  $W(m)$  tends to infinity as  $m$  tends to infinity. Hence there is an integer  $m$  uniquely defined by  $W(m) \leq t^{-1} < W(m+1)$  for any given positive value  $t$ . For this value of  $m$  define  $p$ , where

$$\frac{1}{m+1} < p \leq \frac{1}{m}, \text{ by}$$

$$t^{-1} = \sum_{i=1}^m R(ip). \quad (2.10.1)$$

This gives a unique value for  $p$  in the given range since the right-hand side of (2.10.1) has a negative derivative with respect to  $p$  for

$$p = \frac{1}{m+1},$$

and its second derivative is negative for  $0 < p < 1$ .

Now  $R(F) + R(F+p) + \dots + R(F+mp)$  has a negative second derivative with respect to  $F$ , and for  $F=0$  and  $F=1-mp$  it has the value  $t^{-1}$ ; hence we have

$$R(F) + \dots + R(F+mp) \geq t^{-1} \quad \text{for } 0 \leq F \leq 1-mp. \quad (2.10.2)$$

Similarly

$$R(F) + \dots + R(F+(m-1)p) \geq t^{-1} \quad \text{for } 1-mp \leq F \leq p. \quad (2.10.3)$$

We now suppose that, for a value  $t' > t$  and for all  $x$ ,

$$\int_x^{x+t'w} dF(x) \leq p. \quad (2.10.4)$$

We can write this as

$$\int_F^{F+p} dx \geq t'w. \quad (2.10.5)$$

We can approximate arbitrarily closely to a given distribution by one for which  $f(x) > 0$  and so one for which the distribution function  $F(x)$  has a uniquely defined inverse function. We may therefore suppose that this situation obtains. We may also suppose that the median of the distribution is  $x=0$  (since all conditions are independent of the choice of origin) and we define  $F_1(x)$  by

$$F_1(x) = F(x) \quad \text{for } 0 \leq x,$$

$$F_1(x) = \max(0, F(x+kt'w) - kp) \quad \text{for } x < 0,$$

where  $k$  is the integer such that  $-kt'w \leq x < -(k-1)t'w$ .

We then define  $F_2(x)$  by

$$F_2(x) = F_1(x) \quad \text{for } x \leq 0,$$

$$F_2(x) = \min(1, F_1(x - lt'w) + lp) \quad \text{for } 0 \leq x,$$

where  $l$  is the integer such that  $(l-1)t'w < x \leq lt'w$ . The graph of  $F_2(x)$  thus consists of repetitions of the portion of the graph of  $F(x)$  for  $0 < x < t'w$ , at heights above the  $x$ -axis which ensure that, for  $F_2(x)$ , the inequalities (2.10.4) and (2.10.5) hold, with equality as far as possible.

Using (2.10.5), we have  $F_2(x) = F_1(x) \leq F(x) < \frac{1}{2}$  for  $x < 0$  and  $\frac{1}{2} \leq F(x) = F_1(x) \leq F(x)$  for  $0 \leq x$ . Hence  $R(F_2(x)) \leq R(F(x))$  for all  $x$ , and so

$$w \geq \int_{-\infty}^{\infty} R(F_2(x)) dx. \quad (2.10.6)$$

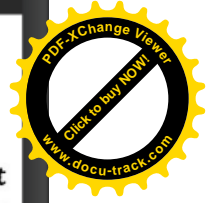
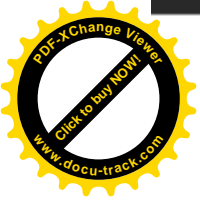
$F_2(x)$  is monotonic increasing and continuous on the right everywhere. It is also continuous on the left, except possibly at points where  $x/t'w$  is an integer. Whenever  $0 < F_2(x) \leq 1 - p$  we have

$$F_2(x + t'w) = F_2(x) + p \quad (2.10.7)$$

and so  $F_2(x)$  defines a distribution with a finite range from  $x_0$  to  $x_1$  where, from (2.10.7),  $mt'w \leq x_1 - x_0 \leq (m+1)t'w$ . If we let  $x_1 - x_0 = mt'w + r$ , then  $F_2(x_0 + r) = F_2(x_1) - mp = 1 - mp$  and  $F_2(x_0 + t'w - 0) = p$ .

Now, using (2.10.7), we have

$$\begin{aligned} \int_{-\infty}^{\infty} R(F_2(x)) dx &= \\ &= \int_{x_0}^{x_0+r} \{R(F_2(x)) + \dots + R(F_2(x) + mp)\} dx \\ &+ \int_{x_0+r}^{x_0+t'w} \{R(F_2(x)) + \dots + R(F_2(x) + (m-1)p)\} dx. \end{aligned}$$



Using (2.10.2), (2.10.3) and (2.10.6), this gives  $w \geq t^{-1} t' w$  so that  $t \geq t'$ . This contradicts the assumption that  $t < t'$  and shows that (2.10.4) must be false. Hence for any  $t' > t$  we have

$$\sup_x \int_x^{x+t'w} dF(x) \geq p$$

and so

$$\sup_x \int_x^{x+tw} dF(x) \geq p,$$

which is the inequality proved by Winsten (1946) who gave tables to assist in the calculation of  $p$  from  $t$ . It is to be noted that unlike the previous inequalities it relates not to probability in a single interval but in a class of intervals.

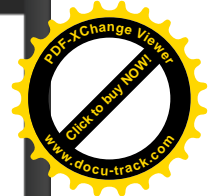
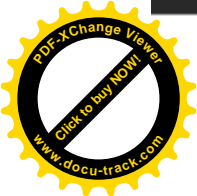
The inequality is best possible, as can be seen from consideration of the discrete distribution with probabilities  $p$  at  $x = \alpha, \alpha + t''w(1 + \epsilon), \dots, \alpha + (m-1)t''w(1 + \epsilon)$  and  $1 - pm$  at  $x = \alpha + mt''w(1 + \epsilon)$ .  $\alpha$  can be chosen so that the median of the distribution is at  $x = 0$ , and we have

$$\sup_x \int_x^{x+t''w} dF(x) = p,$$

while

$$w = t''w(1 + \epsilon) \sum_{i=1}^m R(ip),$$

so that  $t = t''(1 + \epsilon)$ . By taking  $\epsilon$  positive but arbitrarily small we see that no improvement is possible in the inequality.



## CHAPTER III

# UNIVARIATE DISTRIBUTIONS: GEOMETRICAL DATA

### 3.1 Introduction

In the previous chapter the only restriction placed on a p.d.f. was that it should be non-negative, and the distributions giving the values  $L$  and  $U$  were all discrete ones. If we now restrict the p.d.f. or its derivatives in some way, such distributions may become inadmissible and the values of  $L$  and  $U$  may be altered. In this chapter we suppose that the signs of the derivatives are specified, and we show how the method used in Chapter II can be modified to deal with this situation. As examples of the method we shall obtain a number of inequalities obtained by other writers by various methods.

### 3.2 Unimodal distribution: second moment about mode given

We consider first the case of a unimodal distribution for which the second moment about the mode is given and for which  $T$  is an interval symmetrical about the mode; this case is the oldest one considered, the result having been given by Gauss in 1821.

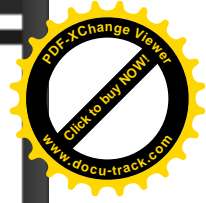
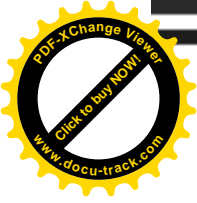
We may fold the distribution about the mode, which we suppose to be at the origin, and we then have

$$\begin{aligned}f'(x) &\leq 0 & \text{for } 0 \leq x \\f(x) &= 0 & \text{for } x < 0,\end{aligned}$$

while  $T$  is  $0 \leq x \leq k$ .

As in Chapter II, we define  $g(x)$  as  $a_0 + a_2 x^2$  and we compare

$$\int_{-\infty}^{\infty} f(x) \chi_T(x) dx \quad \text{with} \quad \int_{-\infty}^{\infty} f(x) g(x) dx.$$



Integration by parts gives for these integrals

$$-\int_{-\infty}^{\infty} f'(x) X_T(x) dx \quad \text{and} \quad -\int_{-\infty}^{\infty} f'(x) G(x) dx \quad \text{respectively,}$$

$$\text{where } X_T(x) = \int_0^x \chi_T(u) du \quad \text{and} \quad G(x) = \int_0^x g(u) du \quad \text{for } 0 \leq x,$$

and  $X_T(x) = G(x) = 0$  for  $x < 0$ . (If  $f(x)$  possesses a finite second moment then it must tend to zero in such a way that  $f(x) X_T(x)$  and  $f(x) G(x)$  tend to zero as  $x$  tends to infinity.) To find  $U$  we want to have

$$\int_{-\infty}^{\infty} f(x) \chi_T(x) dx \leq \int_{-\infty}^{\infty} f(x) g(x) dx$$

and so, since  $f'(x) < 0$ , we need  $X_T(x) \leq G(x)$ ; if, moreover,  $G(x) = X_T(x)$  whenever  $f'(x) \neq 0$  then we shall have strict equality between the integrals and so obtain the exact value of  $U$ .

$$\begin{aligned} \text{For the case considered, } X_T(x) &= 0 \quad \text{for } x \leq 0 \\ &= x \quad \text{for } 0 \leq x \leq k \\ &= k \quad \text{for } k \leq x, \end{aligned}$$

$$\text{and } G(x) = b + a_0 x + a_2 x^3/3 \quad \text{for } 0 \leq x, \quad G(x) = 0 \quad \text{for } x < 0.$$

Since  $f'(x) \neq 0$  at  $x = 0$  we must take  $b = 0$ . (Strictly  $f'(x)$  does not exist at  $x = 0$ , but we can replace all "vertical" parts of the graph of  $f(x)$  by steeply sloping lines and proceed to the limit to get the inequality we want.) To find  $U$  we now have to reduce  $G(x)$  as far as possible while still satisfying  $G(x) \geq X_T(x)$  for  $0 \leq x$ .

We have  $a_2 \geq 0$  (since  $G(x) \geq X_T(x)$  for large positive  $x$ ) and  $a_0 \geq 1$  (since  $G(x) \geq X_T(x)$  for small positive  $x$ ) and so  $G(x) \geq x$ ;  $G(x) = x$  gives  $g(x) = 1$  and  $U = 1$ . This can be seen independently since we can take  $f(x) = (1-\epsilon)/k$  for  $0 \leq x \leq k$ , and  $f(x) = \epsilon/l$  for  $k \leq x \leq k+l$

$$= \sqrt{\left(\frac{3\mu'_2 - k^2}{\epsilon} + \frac{k^2}{4}\right)} - \frac{1}{2}k.$$

This gives the correct value for  $\mu'_2$  and  $P(T) = 1 - \epsilon$ , where  $\epsilon$  can be arbitrarily small. We can describe this process as "adding an

infinite tail of zero probability", corresponding to the process of adding zero probability at infinity when we have no restriction on  $f'(x)$ .

In finding  $L$  we have  $G(x) \leq X_T(x)$  and we want to make  $G(x)$  as large as possible. By increasing  $a_0$  or  $a_2$  we can make the graphs of  $G(x)$  and  $X_T(x)$  touch at  $x = \alpha$  where  $k < \alpha$ . (It would be impossible for the graphs to touch at  $x = \beta$  where  $0 < \beta < k$  since  $a_2 \leq 0$  and  $G(x) \leq a_0 x \leq x$ .)

$$\text{Then} \quad G(x) = k - k(x - \alpha)^2(x + 2\alpha)/2\alpha^3$$

$$\text{giving} \quad g(x) = \frac{3k}{2\alpha^3}(\alpha^2 - x^2)$$

and since we must have  $G'(0) < 1$  we need  $3k < 2\alpha$ . We also have

$$E(g(x)) = (3k/2\alpha) - (3k\mu'_2/2\alpha^3).$$

Hence

$$\frac{d}{d\alpha} E(g(x)) = -\frac{3k}{2\alpha^2} + \frac{9k\mu'_2}{2\alpha^4} = \frac{3k}{2\alpha^4}(3\mu'_2 - \alpha^2).$$

If  $k \geq \sqrt{(4\mu'_2/3)}$  then for  $\alpha > 3k/2$  we have  $3\mu'_2 - \alpha^2 < 0$ , and  $E(g(x))$  is greatest for  $\alpha = 3k/2$ , so that

$$L = 1 - \frac{4\mu'_2}{9k^2}. \quad (3.2.1)$$

If  $k \leq \sqrt{(4\mu'_2/3)}$  then  $E(g(x))$  is greatest for  $\alpha = \sqrt{(3\mu'_2)}$ , giving

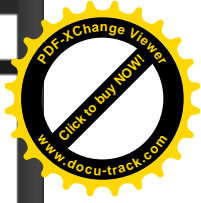
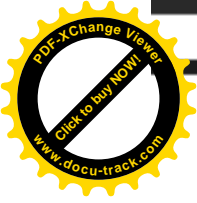
$$L = \frac{k}{\sqrt{(3\mu'_2)}}. \quad (3.2.2)$$

The results (3.2.1) and (3.2.2) are known as the Gauss-Winkler inequalities, having been stated (without proof) by Gauss in 1821 and extended by Winkler in 1866. (See Fréchet (1950).)

### 3.3 Unimodal distribution: first and second absolute moments about mode given

If in addition to  $\mu'_2 = \nu_2$  we also know  $\nu_1$  (i.e.  $\mu'_1$  for the folded distribution) then we take  $G(x)$  to be

$$a_0 x + \frac{a_1 x^2}{2} + \frac{a_2 x^3}{3}.$$



We must have  $3v_2 \geq 4v_1^2$  since

$$\int_0^\infty f'(x) x(u+tx)^2 dx$$

is a form in  $u$  and  $t$  which cannot take positive values.

In evaluating  $U$  we now have the possibility of  $G(x)$  being equal to  $X_T(x)$  at  $x = \alpha$ , where  $0 < \alpha < k$ , or at  $x = \beta$  where  $k < \beta$ . In the latter case, by adding a negative multiple of  $x(x - \beta)^2$  to  $G(x)$  we can make  $G(x)$  equal to  $X_T(x)$  at  $x = k$  also.

In the first case we have

$$3a_1 = -4\alpha a_2, \quad 6(a_0 - 1) = 2\alpha^2 a_2,$$

giving

$$E(g(x)) = \psi = a_2 \left\{ v_2 - \frac{4\alpha v_1}{3} + \frac{\alpha^2}{3} \right\} + 1 \geq 1,$$

and so  $U = 1$ .

In the second case  $G(x) = k + (x - k)(x - \beta)^2/\beta^2$

$$\begin{aligned} \psi &= 3v_2/\beta^2 - 2(2\beta + k)v_1/\beta^2 + (\beta^2 + 2\beta k)/\beta^2 \\ &= 1 + (2k - 4v_1)/\beta + (3v_2 - 2kv_1)/\beta^2 \\ \frac{d\psi}{d\beta} &= (4v_1 - 2k)/\beta^2 - 2(3v_2 - 2kv_1)/\beta^3. \end{aligned}$$

Now if  $k \leq 2v_1$  then  $3v_2 - 2kv_1 \geq 0$  and  $d\psi/d\beta = 0$  for

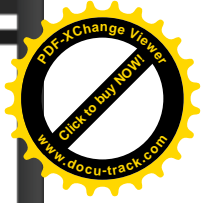
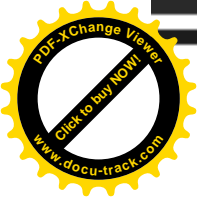
$$\beta = \frac{3v_2 - 2kv_1}{2v_1 - k} > k,$$

giving

$$U = 1 - \frac{(2v_1 - k)^2}{3v_2 - 2kv_1};$$

but if  $2v_1 \leq k$  then  $\psi$  decreases for  $k \leq \beta$  and so  $U = 1$ .

In evaluating  $L$  we can have  $G(x) = X_T(x)$  for  $x = \alpha$  with  $0 < \alpha < k$  and for  $x = \beta$  with  $k < \beta$  or  $G'(0) = X_T'(0) (=1)$  and  $G(\beta) = X_T(\beta)$ , or  $a_2 = 0$ ,  $G(\beta) = X_T(\beta)$ ,  $G'(0) \leq 1$ . (It may be recalled that in Section 2.7 we had an example in which the highest moment produced no effect.)



In the first case we have

$$G(x) = k - k(x - \beta)^2(x + \gamma)/\beta^2\gamma$$

and we need  $k - k(x - \beta)^2(x + \gamma)/\beta^2\gamma = x$  to have roots  $0, \alpha, \alpha$ .

Hence

$$\gamma - 2\beta = -2\alpha \quad \text{and} \quad \beta^2 - 2\beta\gamma + \beta^2\gamma/k = \alpha^2.$$

Hence  $\gamma = 2(\beta - \alpha)$  and  $2\beta^3 - \beta^2(2\alpha + 3k) + 4\alpha\beta k - \alpha^2 k = 0$ ,  
i.e.  $(\beta - \alpha)(2\beta^2 - 3k\beta + k\alpha) = 0$ .

Hence  $\alpha = (3k\beta - 2\beta^2)/k$  and we must have  $\beta < 3k/2$ . Since

$$(k - \alpha) = (k - \beta)(k - 2\beta)/k$$

the condition  $\alpha < k$  is automatically fulfilled.

This expression for  $G(x)$  gives

$$\psi = \psi_1 = \frac{k}{4\beta^3(\beta - k)} \{-3k\nu_2 + 4\nu_1(3k\beta - 2\beta^2) - 9k\beta^2 + 8\beta^3\}.$$

In the second case we have

$$G(x) = k - (x - \beta)^2((2k - \beta)x + \beta k)/\beta^3$$

so that  $\beta \leq 2k$  and

$$\psi = \psi_2 = -\{3\nu_2(2k - \beta) + 2\nu_1(2\beta^2 - 3\beta k) - \beta^3\}/\beta^3.$$

In the third case we have

$$G(x) = k - k(x - \beta)^2/\beta^2$$

and we need  $2k \leq \beta$  and have  $\psi = \psi_3 = -2k(\nu_1 - \beta)/\beta^2$ .

If  $\psi_1 = \psi_2$  then  $(2\beta - 3k)^2(\beta^2 - 4\nu_1\beta + 3\nu_2) = 0$ , and so

$\psi_1 - \psi_2$  is of constant sign for  $k < \beta (< \frac{3k}{2}$  for  $\psi_1$  to be considered).

Now, for  $\beta = k$ , the expression

$$-3k\nu_2 + 4\nu_1(3k\beta - 2\beta^2) - 9k\beta^2 + 8\beta^3$$

equals  $-k^3 + 4k^2\nu_1 - 3k\nu_2$  and is negative unless  $4\nu_1^2 = 3\nu_2$  and  $k = 2\nu_1$ . In this case  $\psi_1 = \nu_1(4\beta^2 - 5\nu_1\beta + 2\nu_1^2)/\beta^3$  (for  $\beta > 2\nu_1$ )

and this tends to 1 as  $\beta$  tends to  $2\nu_1$ , while  $\psi_2$  is 2 for this value of  $\beta$ . Hence in any case  $\psi_1 \leq \psi_2$  and so we calculate  $L$  from  $\psi_1$  for  $k < \beta \leq 3k/2$ , from  $\psi_2$  for  $3k/2 \leq \beta \leq 2k$ , and from  $\psi_3$  for  $2k \leq \beta$ .

Now  $d\psi_1/d\beta = 0$  for

$$-\frac{3}{\beta} - \frac{1}{\beta - k} + \frac{24\beta^2 - 18k\beta - 16\nu_1\beta + 12\nu_1k}{8\beta^3 - 9k\beta^2 + 4\nu_1(3k\beta - 2\beta^2) - 3k\nu_2} = 0,$$

i.e.

$$\frac{4\beta - 3k}{\beta(\beta - k)} = \frac{(4\beta - 3k)(6\beta - 4\nu_1)}{8\beta^3 - 9k\beta^2 + 4\nu_1(3k\beta - 2\beta^2) - 3k\nu_2}$$

i.e.

$$2\beta^3 - \beta^2(4\nu_1 + 3k) + 8k\nu_1\beta - 3k\nu_2 = 0 \quad (3.3.1)$$

i.e.

$$(\beta - k)^2(2\beta + k - 4\nu_1) = -k^3 + 4k^2\nu_1 - 3k\nu_2.$$

Since  $-k^3 + 4k^2\nu_1 - 3k\nu_2 \leq 0$ ,  $d\psi_1/d\beta$  is equal to 0 for just one value of  $\beta$  greater than  $k$ .

$$\text{For } \beta = 3k/2, \frac{d\psi_2}{d\beta} = \frac{-16}{9k^3} (\nu_1 k - \nu_2)$$

and, since  $\psi_1 = \psi_2$  has a double root for  $\beta = 3k/2$ ,  $d\psi_1/d\beta$  has the same value there. Since

$$\frac{d\psi_2}{d\beta} = \frac{(\beta - 3k)(4\nu_1\beta - 6\nu_2)}{\beta^4} \quad \text{and} \quad \frac{d\psi_3}{d\beta} = \frac{2k}{\beta^3} (2\nu_1 - \beta),$$

we have that, for  $k \leq \nu_1$ ,  $\psi_3$  has a maximum at  $\beta = 2\nu_1$ ,  $\psi_2$  is an increasing function for  $3k/2 \leq \beta \leq 2k$ , and  $\psi_1$  is an increasing function for  $k < \beta \leq 3k/2$ .

Hence  $L = k/2\nu_1$ .

For  $\nu_1 \leq k \leq 3\nu_2/4\nu_1$ ,  $\psi_3$  is a decreasing function for  $2k \leq \beta$ , but  $\psi_1$  and  $\psi_2$  are increasing functions in their respective ranges.

Hence  $L = 1 - \nu_1/2k$ .

For  $3\nu_2/4\nu_1 \leq k \leq \nu_2/\nu_1$ ,  $\psi_3$  is a decreasing function and  $\psi_1$  is an increasing function in their respective ranges, but  $\psi_2$  has a maximum for  $\beta = 3\nu_2/2\nu_1$  and

$$L = 1 - \frac{4}{3} \frac{\nu_1^2}{\nu_2} + \frac{8}{9} \frac{k\nu_1^3}{\nu_2^2}.$$

Finally, for  $\nu_2 \leq k\nu_1$ , both  $\psi_2$  and  $\psi_3$  are decreasing functions in their respective ranges, while  $\psi_1$  has a maximum at the value of  $\beta$  given by (3.3.1) and  $L$  is the value obtained by substituting this value of  $\beta$  in  $\psi_1$ .

These results were obtained by Royden (1953) by showing that the problem could be reduced to one in which the graph of the p.d.f. consisted of a bounded number of rectangular blocks.

### 3.4 Unimodal distribution: first and second moments about any point given

In this section we take  $\mu'_1 = 0$ ,  $\mu'_2 = 1$ ,  $T$  as the interval  $x \leq k \leq 0$ , and we suppose that  $f(x)$  has a single maximum, say at  $x = \delta$ .

Since for any choice of  $\gamma$  we must have

$$\int_{-\infty}^{\infty} (x - \delta)(x - \gamma)^2 f'(x) dx \leq 0$$

i.e.  $3 + 2\gamma\delta + \delta^2 \geq 0$  (on integrating by parts), then we must have  $|\delta| \leq \sqrt{3}$ .

In finding  $U$  we need  $G(x) \leq X_T(x)$  for  $x \leq \delta$  and  $G(x) \geq X_T(x)$  for  $\delta \leq x$ , where  $X_T(x) = x - k$  for  $x \leq k$ ,  $X_T(x) = 0$  for  $k \leq x$ .

If  $G(x) = X_T(x)$  at  $x = \alpha$  then, by adding a negative multiple of  $(x - \delta)(x - \alpha)^2$  to  $G(x)$  we can ensure that  $G(x) = X_T(x)$  at some other value  $x = \beta$ . If  $\delta < k$ , then from consideration of the number of times the curve  $y = G(x)$  (a cubic function) can meet the lines  $y = x - k$  or  $y = 0$  we must have  $G(k) = 0$ ,  $G(\alpha) = 0$  ( $k < \alpha$ ), and  $G'(\alpha) = 0$ , unless  $G(x) = x - k$ , so that  $U = 1$ .

This gives

$$G(x) = (x - k)(x - \alpha)^2/(\delta - \alpha)^2$$

and

$$\psi = \psi_0 = (3 + \alpha^2 + 2\alpha k)/(\delta - \alpha)^2.$$

This has a minimum for  $\alpha = -(3 + k\delta)/(k + \delta)$  ( $> k$ ), giving

$$U = (3 - k^2)/(3 - k^2 + (k + \delta)^2) \leq (3 - k^2)/(3 + 3k^2). \quad (3.4.1)$$

If  $k < \delta$  then the graphs of  $G(x)$  and  $X_T(x)$  must touch at  $x = \beta$  ( $\beta < k$ ) and  $x = \alpha$  ( $k < \alpha$ ), so that

$$G(x) = \lambda(x - \delta)(x - \alpha)^2$$

where

$$G(\beta) = \lambda(\beta - \delta)(\beta - \alpha)^2 = \beta - k$$

and

$$G'(\beta) = \lambda(\beta - \alpha)(3\beta - \alpha - 2\delta) = 1.$$

Hence

$$\alpha(k - \delta) + 2\beta^2 - \beta(3k + \delta) + 2k\delta = 0.$$

If we are given  $\delta$  we can eliminate  $\alpha$ , and we find that  $\psi$  is a minimum for a value of  $\beta$  satisfying a cubic equation with coefficients depending on  $\delta$  and  $k$ . We should then have to maximize  $U$  for variation in  $\delta$  to obtain a result valid for any  $\delta$ . It is easier to eliminate  $\delta$  at the outset, and we then have

$$\begin{aligned} \psi &= (3 + \alpha^2 + 2\alpha\delta)/(\beta - \alpha)(3\beta - \alpha - 2\delta) \\ &= (\alpha^3 + \alpha^2\beta + 4\alpha\beta^2 - 6k\alpha\beta + 3\alpha + 3\beta - 6k)/(\alpha - \beta)^3. \end{aligned} \quad (3.4.2)$$

$\alpha$  and  $\beta$  have their ranges of variation restricted by the requirements  $\beta < k$ ,  $k < \alpha$  and  $|\delta| \leq \sqrt{3}$ .

If  $|\delta| = \sqrt{3}$  then the distribution must be rectangular with centre at the origin and range  $2\sqrt{3}$ , so that

$$\left. \begin{aligned} U &= 0 \quad \text{for } k \leq -\sqrt{3} \\ U &= (k + \sqrt{3})/2\sqrt{3} \quad \text{for } -\sqrt{3} \leq k \leq 0. \end{aligned} \right\} \quad (3.4.3)$$

If  $\beta = k$  then  $\delta = k$  and we obtain the result in (3.4.1). If  $\alpha = k$  then, since  $\delta + k > 2\beta$ , we have  $\beta = k$  again.

For stationary values of  $\psi$  inside the region in which  $\alpha$  and  $\beta$  vary we have, from (3.4.2) (denoting the numerator of the expression there by  $E$ ),

$$\frac{3\alpha^2 + 2\alpha\beta + 4\beta^2 - 6k\beta + 3}{E} - \frac{3}{\alpha - \beta} = 0 \quad (3.4.4)$$

and

$$\frac{\alpha^2 + 8\alpha\beta - 6k\alpha + 3}{E} + \frac{3}{\alpha - \beta} = 0. \quad (3.4.5)$$

(3.4.4) gives  $(\alpha + 2\beta - 3k)(2\beta\alpha + \beta^2 + 3) = 0$ . If  $\alpha = -2\beta + 3k$  then (3.4.5) gives  $\beta = (3k^2 + 1)/2k$ , whence  $\alpha = -1/k$ ,  $\delta = -1/k$  also, and  $\psi = 4/9(k^2 + 1)$ . If  $\alpha = -(\beta^2 + 3)/2\beta$  then (3.4.5) gives  $\beta = -1/k$ , which is not possible since  $\beta < k < 0$ , or  $\beta = -\sqrt{3}$ , whence  $\alpha = \sqrt{3}$ ,  $\delta = (6 + 4\sqrt{3}k)/(-2k)$  and  $\psi = (k + \sqrt{3})/(2\sqrt{3})$  as in (3.4.3).

For a bound valid for all  $\delta$  we have

$$U \leq \max \left\{ \frac{3 - k^2}{3 + 3k^2}, \frac{4}{9(k^2 + 1)}, 0, \frac{k + \sqrt{3}}{2\sqrt{3}} \right\} \\ = \max \left\{ \frac{3 - k^2}{3 + 3k^2}, \frac{4}{9(k^2 + 1)} \right\},$$

since  $(k + \sqrt{3})/(2\sqrt{3})$  is a bound only for  $-\sqrt{3} < k < 0$  and then

$$\frac{k + \sqrt{3}}{2\sqrt{3}} < \frac{3 - k^2}{3 + 3k^2}.$$

This gives  $U \leq \frac{3 - k^2}{3 + 3k^2}$  for  $-\sqrt{\frac{5}{3}} \leq k \leq 0$

$$U \leq \frac{4}{9(k^2 + 1)} \quad \text{for } k \leq -\sqrt{\frac{5}{3}}.$$

By taking  $G(x) = 0$  we have  $L = 0$ .

For  $k > 0$  we can obtain values by symmetry.

This result was obtained by Mallows (1956) by a method which he applied to a number of such problems. The method is to determine "extremal distributions" which are such that their distribution functions equal other distribution functions satisfying the same moment conditions at as few points as possible (or at one more than the least possible number), and then to show that among this class of distributions is the one giving the required bound. As we shall see in Section 3.6, it is possible by Mallows's method to deal with a problem intractable by the method used above.

### 3.5 A numerical example

We now take  $\mu'_1 = 0$ ,  $\mu'_2 = 1$ ,  $T$  as  $|x| \leq 2$  and have the conditions that  $f''(x) > 0$  for  $|x| > 1$  and  $f''(x) < 0$  for  $|x| < 1$ . We integrate

$g(x)$  and  $\chi_T(x)$  twice and obtain

$$\begin{aligned} X_T(x) &= -2x - 2 \quad \text{for } x \leq -2, \\ &= \frac{1}{2}x^2 \quad \text{for } |x| \leq 2, \\ &= 2x - 2 \quad \text{for } 2 \leq x \end{aligned}$$

(choosing constants of integration arbitrarily, except that  $X_T(x)$  is to be continuous), and to find  $L$  we require  $G(x) \leq X_T(x)$  for  $|x| \geq 1$  and  $G(x) \geq X_T(x)$  for  $|x| \leq 1$ . If  $G(x)$  satisfies these conditions then so do  $G(-x)$  and  $\frac{1}{2}(G(x) + G(-x))$ , and  $\psi$  is the same in all three cases. Hence we may suppose that  $G(x)$  is an even function of  $x$ . We shall then have  $G(x) = X_T(x)$  at  $x = \pm\alpha$ ,  $x = \pm 1$ , and possibly at  $x = 0$  also. If  $2 \leq \alpha$  then  $G(\alpha) = 2\alpha - 2$ ,  $G'(\alpha) = 2$  and so

$$G(x) = \frac{1}{2} - \frac{(x^2 - 1)\{(2\alpha - 1)(\alpha - 2)x^2 - (6\alpha^4 - 10\alpha^3 - 2\alpha^2 + 5\alpha)\}}{2\alpha(\alpha^2 - 1)^2}$$

Hence

$G(x) = 2x - 2$  for  $x = \alpha$  (double root) and  $x$  such that

$$(2\alpha^2 - 5\alpha + 2)(x + \alpha)^2 = 2(\alpha^2 - 1)^2. \quad (3.5.1)$$

In order that (3.5.1) shall have no root greater than 2 we need

$$(\alpha + 2)^2(2\alpha^2 - 5\alpha + 2) \geq 2(\alpha^2 - 1)^2$$

or

$$\alpha^3 - 2\alpha^2 - 4\alpha + 2 \geq 0$$

i.e.

$$\alpha \geq \alpha_0 = 3.0861 \dots$$

$G(x) = \frac{1}{2}x^2$  for  $x = \pm 1$  and for  $x$  such that

$$(2\alpha - 1)(\alpha - 2)x^2 + (\alpha^3 - 6\alpha^4 + 8\alpha^3 + 2\alpha^2 - 4\alpha) = 0 \quad (3.5.2)$$

i.e.

$$(2\alpha - 1)x^2 + \alpha(\alpha^3 - 4\alpha^2 + 2) = 0.$$

In order that (3.5.2) shall have no real root we need

$$\alpha^3 - 4\alpha^2 + 2 \geq 0$$

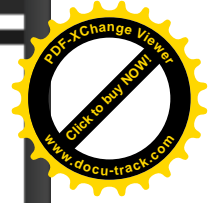
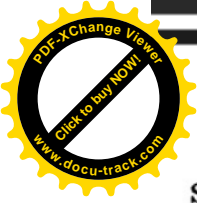
i.e.

$$\alpha \geq \alpha_1 = 3.8662 \dots$$

Now  $\psi = (6\alpha^4 - 10\alpha^3 - 12\alpha^2 + 30\alpha - 10)/\alpha(\alpha^2 - 1)^2$ ,

and has stationary values when

$$3\alpha^6 - 10\alpha^5 - 9\alpha^4 + 50\alpha^3 - 31\alpha^2 + 5 = 0.$$



Since this has no roots with  $\alpha > 3.5$  we have  $\max \psi = \psi(\alpha_1) = .9165 \dots$

For the distribution the graph of whose p.d.f. is the  $x$ -axis for  $|x| \geq \alpha_1$ , the line  $y = 5(x + \alpha_1)/\alpha_1^2(\alpha_1^2 - 1)$  for  $-\alpha_1 \leq x \leq -1$ , the line  $y = (1 - 5/\alpha_1(\alpha_1 + 1)) + x(\alpha_1^2 - 5)/\alpha_1^2$  for  $-1 \leq x \leq 0$ , and which is symmetrical about the  $y$ -axis, we find that

$$\int_{-2}^2 f(x) dx$$

is the same as the value obtained for  $\psi(\alpha_1)$  above, while all other conditions are satisfied if we vary the distribution slightly so as to "round off the corners". Hence  $L = .9165 \dots$

It may be verified that the case  $0 \leq \alpha \leq 2$  leads first to the conclusion that  $\alpha$  must be 0 or 1 and then to the value just given for  $L$ .

When we consider  $U$  we find that the only possible form for  $G(x)$  is  $\frac{1}{2}x^2$  which gives  $U = 1$ . Since there is no term in  $x^4$ , we should expect to obtain  $U = 1$  as the limiting case of a distribution with long tails of small probability, and in fact if  $f(x)$  is symmetrical about  $x = 0$ , with

$$f(x) = \frac{1}{2}(a-1)(a+5)/(a^2+4a+5) \quad \text{for } 0 \leq x \leq 1$$

and

$$f(x) = 20(a-x)^3/(a-1)^4(a^2+4a+5) \quad \text{for } 1 \leq x \leq a,$$

then

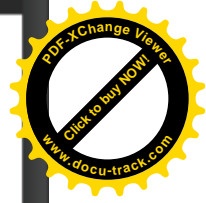
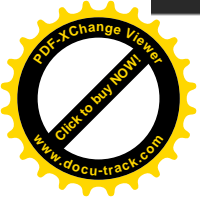
$$\int_{-2}^2 f(x) dx = 1 - 0(a^{-2})$$

and all conditions are satisfied for  $a \geq 5$  (again with the "corners rounded off").

The value  $L = .9165 \dots$  may be compared with the values .75 from Tchebycheff's inequality, using no information about the derivative of  $f(x)$ , and .8889 ... from the Gauss-Winkler inequality using only the unimodality of  $f(x)$ .

### 3.6 Restrictions on the magnitude of $f(x)$ or $F(x)$

In what has gone before we have used the fact that if  $\psi = a_0 + \dots + a_n \mu'_n$  is a support hyperplane to a convex set



defined as the union of half-spaces then we can express  $\psi = a_0 + \dots + a_n \mu'_n$  in the form

$$\sum p_i (a_0 + \dots + a_n x_i^n - \chi_T(x_i)),$$

and this leads to critical distributions which are discrete or else to continuous distributions in which the p.d.f. has a derivative of some order which increases by finite amounts at a finite number of points and is constant in between these points. We can regard a discrete distribution as the limiting case of one with high "peaks" as the height of each peak tends to infinity and its breadth tends to zero; and so, if we restrict the magnitude of  $f(x)$ , such distributions will be excluded.

Instead of expressing  $\psi = a_0 + \dots + a_n \mu'_n$  as we did above, we require an expression for it in the form

$$\int f(x)(a_0 + \dots + a_n x^n - \chi_T(x)) dx;$$

and if  $f(x) \leq \lambda$  and  $\int f(x) dx = 1$ , then  $f(x)$  must be non-zero over an interval of length at least  $1/\lambda$ ; hence  $\psi = a_0 + \dots + a_n \mu'_n$  will now have a non-zero minimum (in the case when we are finding  $L$ ) and we need to estimate this. Instead of finding support hyperplanes to a convex set, we should find hyperplanes that "do not intersect too deeply", and we should expect to find that critical distributions were those in which  $f(x)$  was either 0 or  $\lambda$ . In order to justify this belief it seems easier to abandon the argument which relies on convex sets and to use instead the method of Mallows (1956). To illustrate this method (somewhat simplified since we are dealing with a specific case) we take the example in which  $f(x) \leq \lambda$ , we are given that  $\mu'_1 = 0$ ,  $\mu'_2 = 1$ , and  $T$  is the interval  $x \leq k$ . From a theorem of Achyzer and Krein given in Shohat and Tamarkin (1943) (p. 82) we must have  $\lambda > 1/(2\sqrt{3})$ . (This can be seen otherwise, since for  $\lambda > 1/(2\sqrt{3})$  the most compact distribution is rectangular with range  $> 2\sqrt{3}$  and then  $\mu'_2$  is too large.) If  $k < -1/(2\lambda)$  we note that

$$f(x) = \epsilon(a - |x|)^3 \quad \text{for } 1/(2\lambda) < |x| < a$$

and

$$f(x) = \epsilon(a - |x|)^3 + \lambda(1 - \delta) \quad \text{for } |x| < 1/(2\lambda)$$

satisfies the moment conditions if

$$\epsilon a^6 = 30 \left( 1 - \frac{1 - \delta}{12\lambda^2} \right)$$

$$\epsilon a^4 = 2\delta$$

and  $f(x) \leq \lambda$  if  $\epsilon a^3 \leq \lambda \delta$ .

Also

$$\int_{-\infty}^k f(x) dx = 0(\delta),$$

and for sufficiently small  $\delta$  all conditions will be satisfied for suitable  $\epsilon$  and  $a$ , so that  $L = 0$ . Hence we suppose in what follows that  $k > -1/(2\lambda)$ .

Let  $F^*(x)$  be the distribution function when  $f(x) = \lambda$  for  $k \leq x \leq \alpha$  and  $\beta \leq x \leq \gamma$  ( $\leq k$ ), where  $\alpha, \beta, \gamma$  satisfy

$$\left. \begin{aligned} \alpha - k + \gamma - \beta &= 1/\lambda \\ \alpha^2 - k^2 + \gamma^2 - \beta^2 &= 0 \\ \alpha^3 - k^3 + \gamma^3 - \beta^3 &= 3/\lambda \end{aligned} \right\} \quad (3.6.1)$$

so that the distribution has the prescribed moments. We need to verify that such a distribution exists.

Since  $2(\alpha^3 + \gamma^3) = 3(\alpha + \gamma)(\alpha^2 + \gamma^2) - (\alpha + \gamma)^3$  we obtain from (3.6.1)

$$2(k^3 + \beta^3 + 3/\lambda) = 3(k + \beta + 1/\lambda)(k^2 + \beta^2) - (k + \beta + 1/\lambda)^3$$

i.e.

$$\beta = -(6\lambda^2 + 3k\lambda + 1)/\lambda(6k\lambda + 3);$$

and since  $2\alpha = (\alpha + \gamma) + \sqrt{2(\alpha^2 + \gamma^2) - (\alpha + \gamma)^2}$  we have also

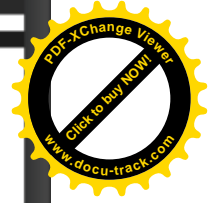
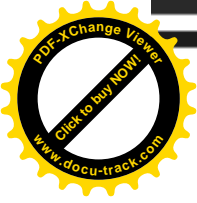
$$\alpha = \frac{1}{2} \left\{ \frac{1}{\lambda} + k + \beta + \sqrt{(k - \beta)^2 - \frac{2}{\lambda}(k + \beta) - \frac{1}{\lambda^2}} \right\},$$

$$\gamma = \frac{1}{2} \left\{ \frac{1}{\lambda} + k + \beta - \sqrt{(k - \beta)^2 - \frac{2}{\lambda}(k + \beta) - \frac{1}{\lambda^2}} \right\}.$$

Since

$$\begin{aligned} (k - \beta)^2 - \frac{2}{\lambda}(k + \beta) - \frac{1}{\lambda^2} &= \\ &= \frac{(6\lambda^2 k^2 + 6\lambda^2 - 2)^2 + (72\lambda^2 - 6)(2\lambda k + 1)}{\lambda^2 (6\lambda k + 3)^2}. \end{aligned}$$

$\alpha$  and  $\gamma$  are real.



Now  $1 + 2\lambda\beta = (1 - 12\lambda^2)/(6k\lambda + 3) \leq 0$  and so

$$(\beta - k + 1/\lambda)^2 \leq (k - \beta)^2 - 2(k + \beta)/\lambda - 1/\lambda^2$$

and we have  $\alpha \geq k$  and also  $k \geq \gamma$ , as we require. Since  $1 + 2\lambda k \geq 0$  we have also

$$(k - \beta + 1/\lambda)^2 \geq (k - \beta)^2 - 2(k + \beta)/\lambda - 1/\lambda^2$$

which gives  $\gamma \geq \beta$ .

Suppose now that for some other distribution  $F(x)$  satisfying the same conditions we have  $F(k) < F^*(k)$ . Then, since the graph of  $F(x)$  nowhere slopes more steeply than the graph of  $F^*(x)$ , the curve  $y = F(x)$  can cut the curve  $y = F^*(x)$  at only one point, say  $x = \theta$ , where  $\beta < \theta < \gamma$ , or else we shall have  $F(x) \leq F^*(x)$  for all  $x$ . In the first case we obtain a contradiction from

$$\begin{aligned} 0 &> \int_{-\infty}^{\infty} (x - \theta)(F(x) - F^*(x)) dx = \\ &\quad - \int_{-\infty}^{\infty} (x - \theta)^2 (F'(x) - F^{*'}(x)) dx = 0 \end{aligned}$$

and in the second case similarly from

$$\int_{\beta}^{\infty} (x - \beta)^2 (F(x) - F^*(x)) dx.$$

Hence  $F(k) \geq F^*(k) = \lambda(\gamma - \beta)$  with  $\beta, \gamma$  as given above, and since  $F^*(x)$  is a possible distribution function this gives

$$L = \lambda(\gamma - \beta).$$

As  $\lambda$  tends to infinity we find that  $\beta$  tends to  $-1/k$ ,  $\alpha$  to  $k$ , and  $\gamma$  to  $-1/k$ , while  $\lambda(\gamma - \beta)$  is asymptotically  $k/(k - \beta)$  which tends to  $k^2/(k^2 + 1)$  as in Section 2.5.

If  $\lambda = (2\pi)^{-1}$  (the value for the normal distribution with unit variance) and  $k = 2$ , then we have  $L = .8777 \dots$ , in contrast to  $L = .8$  when we impose no upper bound on  $f(x)$ .

Another type of restriction on the distribution function which was considered by von Mises (1938) is as follows. Let  $x$  be a non-negative variable and let the graph of  $y = F(x)$  lie below the straight

line joining  $(x_1, F(x_1))$  and  $(z, 1)$  for all  $x \geq x_0$  (where  $x_0 < x_1$ ) and for some value  $z (> x_1)^{(*)}$ . Also let the absolute moment  $\nu_r$  be given. The condition on the graph of  $y = F(x)$  can be written as

$$F(x) \leq H(x) = F(x_1) + (x - x_1)(1 - F(x_1))/(z - x_1). \quad (3.6.2)$$

Then 
$$\nu_r \geq \int_{x_0}^z x^r dF(x) \geq \int_{x_0}^z x^r dH(x)$$

(because any value of  $F$  corresponds to a larger value of  $x$  than does the same value of  $H$ )

i.e. 
$$\nu_r \geq \frac{z^{r+1} - x_0^{r+1}}{r+1} \cdot \frac{1 - F(x_1)}{z - x_1}, \quad (3.6.3)$$

Since the right-hand side of (3.6.3) tends to infinity as  $z$  tends to infinity or to  $x_1 + 0$  it must have a minimum value for some value  $\zeta$  where  $x_1 < \zeta$ , and this is given by

$$(r+1)\zeta^r/(\zeta^{r+1} - x_0^{r+1}) = 1/(\zeta - x_1)$$

or 
$$r\zeta^{r+1} - (r+1)x_1\zeta^r + x_0^{r+1} = 0. \quad (3.6.4)$$

By differentiating once more with respect to  $\zeta$ , it can be seen that the equation (3.6.4) defines  $\zeta (> x_1)$  uniquely. We now have, independently of  $z$ ,

$$\nu_r \geq \frac{\zeta^{r+1} - x_0^{r+1}}{\zeta - x_1} \cdot \frac{1 - F(x_1)}{r+1}$$

or, using (3.6.4),

$$\nu^r > (1 - F(x_1)) \zeta^r$$

i.e. 
$$F(x_1) > 1 - \nu^r/\zeta^r. \quad (3.6.5)$$

Now from (3.6.2) we have

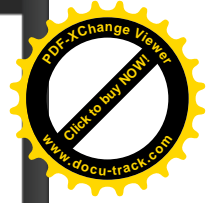
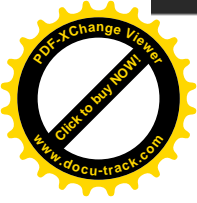
$$0 \leq F(x_0) \leq F(x_1) + (x_0 - x_1)(1 - F(x_1))/(z - x_1)$$

and this gives

$$z \geq Z = x_0 + (x_1 - x_0)/F(x_1), \quad (3.6.6)$$

---

(\*) This means that the graph of the distribution function lies beneath one of its tangents at all points to the right of some given point.



so that if  $\zeta < Z$  we can improve on the inequality (3.6.5). Using the value  $Z$  for  $z$  in (3.6.3) we then have

$$\begin{aligned} \nu_r &\geq (Z^{r+1} - x_0^{r+1})(1 - F(x_1))/(Z - x_1)(r + 1) \\ &= (Z^{r+1} - x_0^{r+1})/(Z - x_0)(r + 1) \quad (\text{using (3.6.6)}) \\ &= (Z^r + Z^{r-1}x_0 + \dots + x_0^r)/(r + 1) \geq x_0^r. \end{aligned}$$

If  $\tau$  is defined by  $\nu_r = (\tau^{r+1} - x_0^{r+1})/(\tau - x_0)(r + 1)$  then  $Z \leq \tau$  and

$$F(x_1) = (x_1 - x_0)/(Z - x_0) \geq (x_1 - x_0)/(\tau - x_0). \quad (3.6.7)$$

The inequality (3.6.7) holds with equality if the distribution is rectangular with range from  $x_0$  to  $z$ ; consideration of the inequalities which lead to (3.6.5) shows that the same distribution must be used to achieve equality there, but comparison of (3.6.7) and (3.6.5) shows that the latter inequality is true only for the special value  $x_1 = \tau = Z$ .

For the normal distribution with unit variance, taking  $r = 2$ , we have  $\nu_2 = 1$ , and if  $x_1 = 2$ , we can take  $x_0$  as any value not less than 0. Taking  $x_0$  as 0, its least possible value, we have  $\zeta = 3$  and  $\tau = \sqrt{3}$  so that  $\zeta > \tau$ , and we use (3.6.5) to give  $F(2) \geq 1 - \frac{1}{9}$  (the value given by the Gauss-Winkler inequality). If  $x_1 = 1$  then  $\zeta = \frac{3}{2} < \tau$ , and we obtain from (3.6.5)  $F(1) \geq 1 - \frac{4}{9}$ , but from (3.6.7) we obtain  $F(1) \geq \frac{1}{3}$  which is a better value.

### 3.7 Mean range given for unimodal symmetrical distribution

As in the case of Chapter 2 we end the present chapter with an inequality in terms of mean range  $w$  from a sample of given size  $n$ . We suppose that the distribution is symmetrical and unimodal (with mode at the origin) and find a lower bound for the case when  $T$  is  $|x| \leq \lambda w$ . Although — in contrast to the earlier example of the use of mean range — we obtain an inequality for a single interval, we cannot use the method employed in most examples because the mean range is expressible only as the expectation of a function of the distribution function; instead we use a special argument.

Let  $f(\lambda w) = h$ ; then for  $\lambda w < x$  we have

$$F(x) \leq F(\lambda w) + h(x - \lambda w)$$

from the unimodal property of the distribution. Also, for  $0 < x < \lambda w$ , we have

$$F(x) \leq F(\lambda w) - h(\lambda w - x).$$

Hence, for  $0 < x$ ,

$$\frac{1}{2} < F(x) \leq F_1(x) = F(\lambda w) - h(\lambda w - x). \quad (3.7.1)$$

Hence

$$w \geq 2 \int_0^{x_1} (1 - (F_1(x))^n - (1 - F_1(x))^n) dx$$

where

$$hx_1 = 1 - F(\lambda w) + h\lambda w,$$

i.e.  $F_1(x) = 1$  for  $x = x_1$ ,  $F_1(x) < 1$  for  $x < x_1$ .

Hence

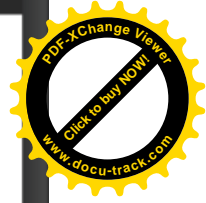
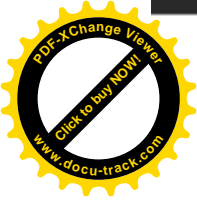
$$\begin{aligned} \frac{1}{2}w &\geq x_1 - \frac{1 - (F(\lambda w) - h\lambda w)^{n+1}}{(n+1)h} - \frac{(1 - F(\lambda w) + h\lambda w)^{n+1}}{(n+1)h} \\ &= \frac{1 - F(\lambda w) + h\lambda w}{h} - \frac{1}{h(n+1)} + \frac{(F(\lambda w) - h\lambda w)^{n+1}}{(n+1)h} \\ &\quad - \frac{(1 - F(\lambda w) + h\lambda w)^{n+1}}{(n+1)h}. \end{aligned}$$

Putting  $h\lambda w = t$  and  $F(\lambda w) - F(-\lambda w) = 2F(\lambda w) - 1 = P$ , we have

$$\begin{aligned} \frac{1}{\lambda} &\geq \frac{1 - P}{t} + 2 - \frac{2}{(n+1)t} + \frac{(P+1-2t)^{n+1}}{2^n(n+1)t} \\ &\quad - \frac{(1-P+2t)^{n+1}}{2^n(n+1)t} \quad (3.7.2) \\ &= Q(t), \text{ say.} \end{aligned}$$

We minimize  $Q(t)$  with respect to  $t$  and interpret the resulting inequality as an inequality for  $P$  in terms of  $\lambda$ .

$$\begin{aligned} \text{Now } t^2 Q'(t) &= -(1-P) + \frac{2}{n+1} \\ &\quad - \frac{(P+1-2t)^{n+1}}{2^n(n+1)} - \frac{(P+1-2t)^n t}{2^{n-1}} \\ &\quad + \frac{(1-P+2t)^{n+1}}{2^n(n+1)} - \frac{(1-P+2t)^n t}{2^{n-1}} \end{aligned}$$



and

$$\frac{d}{dt}(t^2 Q'(t)) = \frac{nt}{2^{n-2}} \{(1+P-2t)^{n-1} - (1-P+2t)^{n-1}\} \geq 0$$

since, from (3.7.1), we have  $t \leq \frac{1}{2}P$ . When  $t = 0$ ,

$$t^2 Q'(t) = -(1-P) + \frac{2}{n+1} - \frac{(P+1)^{n+1}}{2^n(n+1)} + \frac{(1-P)^{n+1}}{2^n(n+1)}.$$

This is zero when  $p=1$  and its derivative with respect to  $P$  is

$$1 - 2^{-n}(P+1)^n - 2^{-n}(1-P)^n > 0 \quad \text{for } 0 \leq P < 1,$$

so that  $t^2 Q'(t) < 0$  when  $t = 0$  and increases steadily with  $t$ . Now

$$\frac{1}{4}P^2 Q'(\frac{1}{2}P) = -1 + P + \frac{2}{n+1} - \frac{P}{2^{n-1}}$$

and if this is negative, i.e. if

$$P < \frac{n-1}{n+1} \cdot \frac{2^{n-1}}{2^{n-1}-1},$$

then  $Q(t)$  takes its least value for  $t = \frac{1}{2}P$  (the greatest permissible value of  $t$ ) and we have

$$\frac{1}{\lambda} \geq \frac{2(n-1)}{(n+1)P}. \quad (3.7.3)$$

If, however,

$$P \geq \frac{n-1}{1+n} \cdot \frac{2^{n-1}}{2^{n-1}-1}$$

then  $Q'(t)$  vanishes for just one value of  $t$  in the interval  $0 < x \leq \frac{1}{2}P$  and for this value (3.7.2) gives the inequality for  $P$  in terms of  $\lambda$ .

To deal with the equations easily put  $1-P+2t = 2y$  to give

$$0 = 2t - 2y + \frac{2}{n+1} - \frac{2}{n+1}(1-y)^{n+1} - 2t(1-y)^n + \frac{2}{n+1}y^{n+1} - 2ty^n$$

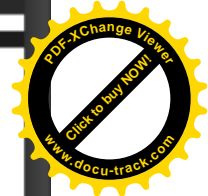
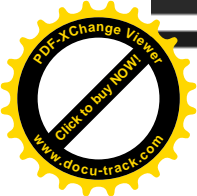
$$\text{i.e.} \quad t = \frac{y - \frac{1}{n+1} + \frac{(1-y)^{n+1}}{n+1} - \frac{y^{n+1}}{n+1}}{1 - y^n - (1-y)^n} \quad (3.7.4)$$

and

$$\begin{aligned} \frac{1}{\lambda} &\geq \frac{2y}{t} - \frac{2}{(n+1)t} + \frac{2(1-y)^{n+1}}{(n+1)t} - \frac{2y^{n+1}}{(n+1)t} \\ &= 2(1-y^n - (1-y)^n). \end{aligned} \quad (3.7.5)$$

For a given  $\lambda$ , (3.7.5) gives bounds on  $y$  from which (3.6.4) gives bounds on  $t$  and hence on  $P$ . From the method of proof there exists a distribution for which the bounds are exact.

If with  $n = 2$  we take the value  $w = 2\pi^{-1/2}$  (which is the correct value for the normal distribution with unit variance) and  $\lambda = \sqrt{\pi}$ , then if  $P < \frac{2}{3}$  we have  $P \geq 2\lambda/3$  which is impossible. Hence  $\frac{2}{3} < P$  and, from (3.7.5), we have  $y(1-y) \leq 1/(4\sqrt{\pi})$ . Also, since  $2y = 1 - P + 2t$ , we have  $0 \leq y \leq \frac{1}{2}$  and so  $0 \leq y \leq .1699 \dots$ . Hence, since  $1 - P = 2(y - t) = (3y - 4y^2)/3(1 - y)$ , we have  $1 - P \leq .1583 \dots$  or  $P \geq .8417 \dots$ . The bound obtained is thus less good than the one given by the Gauss-Winkler inequality, but it improves as we use the mean range from larger samples. Winsten, who first obtained the above inequality (1946), gives tables to assist in the computation of the bound for  $P$ .



## CHAPTER IV

# MULTIVARIATE DISTRIBUTIONS

### 4.1 Introduction

In this chapter we consider the extension of Tchebychef's inequality to  $n$ -variate distributions ( $n \geq 2$ ). As might be expected, the increase in the number of dimensions is accompanied by an increase in the complexity of the working and of the results, and although quite a lot of recent work has been on this aspect of the problem, there is as yet not the completeness or generality which we found in Chapter II. The regions which will be considered in the space of the variables are simpler (being in fact all connected sets, in contrast to the example in Section 2.4), the conditions assumed are simpler (in most cases involving moments of order no higher than the second), yet even when a general method exists it leads to an algebraic problem for which no method of solution has yet been found. Moreover, fewer geometrical restrictions have been employed with multivariate distributions. What does carry over from the earlier work, however, is the general idea of studying a function whose expectation involves the given expectations and using properties of convexity.

### 4.2 Second-order moments: rectangular region

In this section we shall assume that we are given the second-order moments of the distribution (with all means equal to zero) and that  $T$  is the rectangular region  $|x_i| \leq d_i$  ( $i = 1, \dots, n$ ).

By introducing zero probability at infinity we shall have  $U = 1$ .

Without loss of generality we assume that  $d_i = 1$  ( $i = 1, \dots, n$ ) since we can scale the  $x_i$ ; we can also apply a more general transformation to the  $x_i$  to deal with the case of a parallelepipedal region, but we shall do this explicitly in only one example (see Section 4.6).

Let the given moments be  $E(x_i x_j) = \mu_{ij}$  and let  $M$  denote the matrix  $\{\mu_{ij}\}$ .

The function corresponding to that used in Chapter II is  $g(x_1, \dots, x_n) = a_0 + x\alpha' + xAx'$ , where  $x$  is the row vector  $(x_1, \dots, x_n)$ ,  $\alpha$  is a row vector, and  $A$  is a symmetric matrix.

We require  $a_0 + x\alpha' + xAx' \leq 1$  for all  $x$  and

$$a_0 + x\alpha' + xAx' \leq 0,$$

except when  $|x_i| \leq 1$  ( $i = 1, \dots, n$ ). If  $L > 0$  then  $0 < a_0 \leq 1$  and  $A$  must be negative definite. If  $g(x_1, \dots, x_n) = 1$  when some  $x_i$ , say  $x_I$ , equals 1, we can add a positive multiple of  $(x_I - 1)^2$  to increase  $a_0$ , while if  $a_0 = 1$  we can add a positive multiple of  $x_I^2$  to make  $g(x_1, \dots, x_n) = 1$  when  $x_I = 1$ . Finally we may assume that  $g(x_1, \dots, x_n)$  is symmetrical in the  $x_i$  so that  $\alpha = 0$ . Hence we take  $g(x_1, \dots, x_n)$  as  $1 + xAx'$  and we have  $\psi = 1 + \text{tr } AM$  (where "tr" denotes the trace of the matrix).

We now have to maximize  $\psi$  with respect to  $A$ , subject to  $A$  being negative definite and such that

$$x'Ax \leq -1 \quad (4.2.1)$$

whenever any one of  $x_1, \dots, x_n$  is numerically greater than unity, equality obtaining in at least one instance.

We now find the conditions which  $A$  satisfies when this maximum is attained.

First suppose that 
$$A = \begin{pmatrix} a_{11} & a \\ a' & A_{22} \end{pmatrix},$$

where  $a$  is the row vector  $(a_{12}, \dots, a_{1n})$ , and that

$$A^{-1} = B = \begin{pmatrix} b_{11} & b \\ b' & B_{22} \end{pmatrix}.$$

Since  $b_{11}a + bA_{22} = (0, \dots, 0)$  we have

$$b = -b_{11}aA_{22}^{-1}. \quad (4.2.2)$$

Since  $b_{11}a_{11} + ba' = 1$  we have, from (4.2.2),

$$b_{11}(a_{11} - aA_{22}^{-1}a') = 1. \quad (4.2.3)$$

By multiplying the matrices in the reverse order we can interchange the  $a$ 's and  $b$ 's in (4.2.2) and (4.2.3) to give

$$a = -a_{11} b B_{22}^{-1} \quad (4.2.4)$$

and 
$$a_{11} (b_{11} - b B_{22}^{-1} b') = 1. \quad (4.2.5)$$

Writing  $y$  for  $(x_2, \dots, x_n)$ , we have

$$\begin{aligned} (x_1, y) A (x_1, y)' &= a_{11} x_1^2 + 2y a' x_1 + y A_{22} y' \\ &= b_{11}^{-1} x_1^2 + (y - b_{11}^{-1} b x_1) A_{22} (y - b_{11}^{-1} b x_1)', \end{aligned}$$

using (4.2.2) and (4.2.3). Since  $A_{22}$  is negative definite this gives  $x A x' \leq b_{11}^{-1} x_1^2$  with equality only for  $b_{11} y = b x_1$ . Hence, using (4.2.1), we have  $b_{11}^{-1} \leq -1$  and similarly  $b_{22}^{-1} \leq -1, \dots, b_{nn}^{-1} \leq -1$ , with equality in at least one instance.

If  $b_{11} < -1$  let

$$B(\delta) = \begin{pmatrix} b_{11} + \delta & b \\ b' & B_{22} \end{pmatrix}.$$

Since  $B$  is negative definite, so, for sufficiently small  $\delta$ , is  $B(\delta)$  and so is  $\{B(\delta)\}^{-1} = A(\delta)$ .

By direct evaluation of  $\{B(\delta)\}^{-1}$  the element in its first row and column, say  $a_{11}(\delta)$ , is  $a_{11}/(1 + a_{11} \delta)$ . Using (4.2.4) and (4.2.5), we have that

$$a_{11}(\delta) \begin{pmatrix} 1 & -b B_{22}^{-1} \\ -B_{22}^{-1} b' & B_{22}^{-1} b' b B_{22}^{-1} \end{pmatrix} B(\delta) + \begin{pmatrix} 0 & 0 \\ 0 & B_{22}^{-1} \end{pmatrix} B(\delta) = I$$

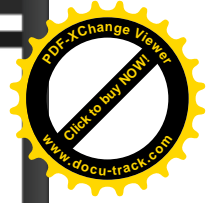
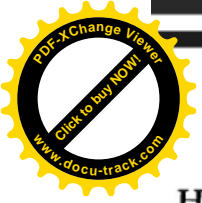
whence

$$A(\delta) = a_{11}(\delta) \begin{pmatrix} 1 & -b B_{22}^{-1} \\ -B_{22}^{-1} b' & B_{22}^{-1} b' b B_{22}^{-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & B_{22}^{-1} \end{pmatrix}.$$

Now, for some  $\theta$  such that  $0 < \theta < 1$ , we can find arbitrarily small  $\delta_1, \delta_2$  such that

$$\frac{\theta \delta_1}{1 + a_{11} \delta_1} + \frac{(1 - \theta) \delta_2}{1 + a_{11} \delta_2} = 0$$

and this is the same as  $\theta a_{11}(\delta_1) + (1 - \theta) a_{11}(\delta_2) = a_{11}$ .



Hence

$$\theta A(\delta_1) + (1 - \theta) A(\delta_2) = A,$$

and  $A$  is an interior point of the set of possible matrices. Hence  $\psi = 1 + \text{tr } AM$ , which is linear in the elements of  $A$ , cannot achieve its maximum at  $A$ . Hence, for maximum  $\psi$ , we must have

$$b_{ii} = -1 \quad \text{for } i = 1, \dots, n. \quad (4.2.6)$$

The restriction that  $B$  be negative definite means that its characteristic roots are negative; the limiting case is when a characteristic root of  $B$  tends to zero and consequently a characteristic root of  $B^{-1}$  tends to minus infinity. This would make  $\text{tr } B^{-1} M$  tend to minus infinity, and hence the maximum of  $\text{tr } B^{-1} M$  occurs at an interior point of the set of matrices  $B$ . If we consider the matrices  $B$  which are linear in some variable  $t$  then we have

$$\frac{d^2}{dt^2} (\text{tr } B^{-1} M) = \frac{d}{dt} \left( \frac{d}{dt} \text{tr } B^{-1} M \right) = \frac{d}{dt} \left( -\text{tr } B^{-1} \frac{dB}{dt} B^{-1} M \right)$$

$$(\text{since } B \frac{d}{dt} (B^{-1}) + \frac{dB}{dt} B^{-1} = 0)$$

$$= 2 \text{tr } B^{-1} \frac{dB}{dt} B^{-1} \frac{dB}{dt} B^{-1} M$$

which is non-positive since  $B^{-1}$  is negative definite and  $M$  is positive definite. Hence the maximum with respect to  $t$  of  $\text{tr } B^{-1} M$  is unique; let it be attained when  $B = C$ .

Consider in particular the case when  $t$  is  $b_{ij}$ , an element of  $B$  off the main diagonal. Since for any matrix  $G$ ,  $BGB^{-1}$  and  $G$  have the same characteristic roots and hence the same trace, we have that

$$\text{tr } B^{-1} \frac{dB}{db_{ij}} B^{-1} M = \text{tr } \frac{dB}{db_{ij}} B^{-1} MB^{-1} = \angle d_{ij},$$

where  $D = B^{-1} MB^{-1}$ . But at the maximum

$$\frac{d}{db_{ij}} (\text{tr } B^{-1} M) = -\text{tr } B^{-1} \frac{dB}{db_{ij}} B^{-1} M = 0,$$

so that  $C^{-1}MC^{-1}$  is a diagonal matrix, say with elements  $\gamma_1, \dots, \gamma_n$  (which are positive, since  $M$  is positive definite). Now  $C^{-1}MC^{-1}C$  has diagonal elements  $-\gamma_1, \dots, -\gamma_n$  and therefore  $\text{tr } C^{-1}M = -(\gamma_1 + \dots + \gamma_n)$  so that  $L = 1 - (\gamma_1 + \dots + \gamma_n)$  (unless  $L$  is 0).

$L$  is attained by discrete distributions as follows. If  $\gamma_1 + \dots + \gamma_n \leq 1$ , let  $\frac{1}{2}\gamma_i$  be the probability at the point  $(c_{i1}, \dots, c_{in})$  and its negative, and let there be probability  $1 - (\gamma_1 + \dots + \gamma_n)$  at the origin. Since  $c_{ii} = -1$ , the probabilities  $\frac{1}{2}\gamma_i$  can be counted as falling outside  $T$ , and we have  $P(T) = 1 - (\gamma_1 + \dots + \gamma_n)$ ; we have means zero and also  $E(x'x) = CC^{-1}MC^{-1}C = M$ , as required.

If  $1 < \gamma_1 + \dots + \gamma_n$  then we take probabilities  $\gamma_i/2\sqrt{\sum \gamma_i}$  at  $\pm\sqrt{(\sum \gamma_i)}(c_{i1}, \dots, c_{in})$ . This gives the correct covariance matrix, and at least one coordinate is greater than unity in absolute value.

We now consider how to solve for  $B$ , and hence  $\gamma_1, \dots, \gamma_n$ , for a given matrix  $M$ .

If  $n = 2$  then  $B$  is  $\begin{pmatrix} -1 & \beta \\ \beta & -1 \end{pmatrix}$  with  $|\beta| < 1$  for negative definiteness, and, if  $M$  is  $\begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$ ,

we require  $\begin{pmatrix} -1 & -\beta \\ -\beta & -1 \end{pmatrix} \begin{pmatrix} \sigma_1^2\rho & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix} \begin{pmatrix} -1 & -\beta \\ -\beta & -1 \end{pmatrix}$

to be diagonal, which leads to

$$\sigma_1\sigma_2\rho\beta^2 + (\sigma_1^2 + \sigma_2^2)\beta + \sigma_1\sigma_2\rho = 0.$$

The product of the roots of this quadratic in  $\beta$  is 1, and the sum has sign opposite to that of  $\rho$ . Hence

$$\beta = \frac{-(\sigma_1^2 + \sigma_2^2) + \sqrt{((\sigma_1^2 + \sigma_2^2)^2 - 4\sigma_1^2\sigma_2^2\rho^2)}}{2\sigma_1\sigma_2\rho}.$$

$$\text{tr } B^{-1}M = -(\sigma_1^2 + 2\beta\sigma_1\sigma_2 + \sigma_2^2)/(1 - \beta^2),$$

and this is

$$-\frac{1}{2}(\sigma_1^2 + \sigma_2^2 + \sqrt{((\sigma_1^2 + \sigma_2^2)^2 - 4\sigma_1^2\sigma_2^2\rho^2)}).$$

Hence  $L$  is

$$1 - \frac{1}{2}(\sigma_1^2 + \sigma_2^2 + \sqrt{((\sigma_1^2 + \sigma_2^2)^2 - 4\sigma_1^2\sigma_2^2\rho^2)}).$$

This is equivalent to a result due to Lal (1955); the special case  $\sigma_1 = \sigma_2$  was dealt with earlier by Berge (1937).

For larger values of  $n$  no general method (other than solving by successive approximation the equations which arise) is known for finding  $B$  and  $A$ , and this presents one of the most interesting unsolved problems of the subject.

The theory above is due to Olkin and Pratt (1958); the same problem was solved in similar terms by Whittle (1958b), independently but slightly later.

An explicit result using only some covariances was given by Birnbaum and Marshall (1961). It is assumed that not more than  $(n-1)$  values of  $\mu_{ij}$  are known, including, for given  $j$ , not more than one value with  $i < j$ . The result is not, in general, best possible.

#### 4.3 Second-order moments: region similar to an orthant

If in the last example we replace  $|x_i| \leq d_i$  by  $x \leq d_i$ , then  $T$  is similar to an orthant (i.e. the generalization to  $n$  dimensions of a quadrant in two dimensions and an octant in three).

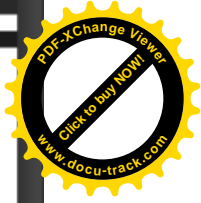
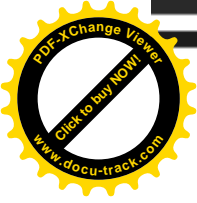
As before, we take  $d_i = 1$  ( $i = 1, \dots, n$ ) and to find  $L$  we consider  $g = a_0 + x\alpha' + xAx'$ , where now

$$a_0 + x\alpha' + xAx' \leq 1 \quad \text{for all } x$$

and  $a_0 + x\alpha' + xAx' \leq 0$  except when  $x_i \leq 1$  ( $i = 1, \dots, n$ ).

We can again take  $A$  to be negative definite, but we can no longer appeal to symmetry to make  $\alpha = (0, \dots, 0)$  nor  $a_0 = 1$ . However, if all the coefficients of correlation between the  $x$ 's are equal and all variances are equal, we can use symmetry in a different way to ensure that all the components of  $\alpha$  are equal, and write  $a_0 + x\alpha' + xAx'$  as  $(x - ke) A (x - ke)'$ , where  $e$  is  $(1, \dots, 1)$ . (To do this we take  $(n!)^{-1} \sum g$  over all permutations of  $x_1, \dots, x_n$ .)

The theory developed in Section 4.2 still applies, and we can show that  $A^{-1}$  must be negative definite with diagonal terms all  $-1$ . This was done by Marshall and Olkin (1960a), and their result is (in our terminology)



$$L = \max (0,$$

$$1 - \frac{n\sigma^2 \{ \sqrt{[(1 + (n-1)\rho)(1 + \sigma^2 - \sigma^2(n-1)(1-\rho))]} + (n-1)\sqrt{(1-\rho)^2} \}}{\{n + \sigma^2(1 + (n-1)\rho)\}^2})$$

where  $\sigma^2 = E(x_i^2)$ ,  $\rho\sigma^2 = E(x_i x_j)$  ( $i \neq j$ ).

The case of unequal covariances has not been considered, and although particular cases could be dealt with numerically by maximizing  $\psi = E(g)$ , it seems likely that a general expression would be complicated.

#### 4.4 A continuous set of variables

The ideas of the previous section have been generalized to the case of a continuous set of variables by Whittle (1958a).

We let  $E(x(s)x(t)) = v(s, t)$ , where  $0 \leq s, t \leq 1$ ,  $E(x(t)) = 0$ ,  $v(t, t) = \phi^2(t)$ , and

$$\left( \frac{\partial^2 v}{\partial s \partial t} \right)_{s=t} = \psi^2(t),$$

and we suppose that

$$E(x(t)^2) \text{ is finite for } 0 \leq t \leq 1 \quad (4.4.1)$$

and that  $x'(t)$  exists and

$$E(x'(t)^2) \text{ is finite for } 0 \leq t \leq 1. \quad (4.4.2)$$

Now consider the functional

$$\begin{aligned} S(x, y) \\ = \frac{x(0)y(0) + x(1)y(1)}{2\alpha^2} + \frac{1}{2\theta\alpha^2} \int_0^1 \{ \theta^2 x(t)y(t) + x'(t)y'(t) \} dt, \end{aligned}$$

and let  $P$  be the probability that  $|x(t)| < \alpha$  in  $(0, 1)$ .

We need to be able to invert the order of the limit operations involved in taking expectations, integrating over  $t$ , and differentiating  $x$  with respect to  $t$ ; conditions (4.4.1) and (4.4.2) are sufficient for this. (See, e.g., Loève (1955), Section 7.2, for the relevant theory.)

We then have

$$\begin{aligned} E(s(x, x)) &= \frac{\phi^2(0) + \phi^2(1)}{2\alpha^2} + \frac{1}{2\theta\alpha^2} \int_0^1 (\theta^2 \phi^2(t) + \psi^2(t)) dt \\ &\leq \frac{\phi^2(0) + \phi^2(1)}{2\alpha^2} + \frac{1}{\alpha^2} \left( \int_0^1 \phi^2(t) dt \cdot \int_0^1 \psi^2(t) dt \right)^{\frac{1}{2}}. \end{aligned}$$

If  $y_s(t) = \alpha^2 e^{-\theta|t-s|}$  it may be verified, by integration by parts, that  $S(x, y_s(t)) = x(s)$ , while  $S(y_s(t), y_s(t)) = \alpha^2$ .

Using Schwarz's inequality, we have

$$S(x(t), x(t)) \geq (x(t))^2 / \alpha^2$$

and so  $E(S(x, x)) \geq 1 - P$ , whence we obtain a bound for  $P$ .

Whittle discusses the conditions on the behaviour of  $x$  to ensure the existence of a function such as  $y_s(t)$  above.

#### 4.5 Variances given: rectangular regions

As another variation on the problem in Section 4.2 we may suppose that covariances are not given, so that  $g$  is now of the form  $a_0 + x\alpha' + xAx'$  with  $A$  diagonal. As before, we can take  $\alpha = (0, \dots, 0)$  and  $a_0 = 1$ , so that  $g$  becomes  $1 + \sum a_{ii} x_i^2$ . Every  $a_{ii}$  must be less than or equal to  $-1$ , and the maximum of  $g$  is obviously attained by making the  $a_{ii}$  all  $-1$ , giving

$$L = \max(0, 1 - \sum \sigma_i^2). \quad (4.5.1)$$

If  $\sum \sigma_i^2 < 1$  then we can obtain this value of  $L$  with the discrete distribution for which  $\Pr(0, \dots, 0) = 1 - \sum \sigma_i^2$ ,  $\Pr(\pm 1, 0, \dots, 0) = \frac{1}{2} \sigma_i^2$ , etc. It will be noticed that in this distribution we have all covariances zero, but that the distributions of  $x_1, \dots, x_n$  are not independent. If the distributions were known to be independent then the repeated application of Tchebychef's inequality in one dimension would give the stronger bound  $L = (1 - \sigma_1^2) \dots (1 - \sigma_n^2)$ .

In the same way that the method for a finite rectangular region was extended in Section 4.3 to a region similar to an orthant, so the

inequality (4.5.1) can be extended to the same region to give

$$L = \max \left( 0, 1 - \sum \frac{\sigma_i^2}{1 + \sigma_i^2} \right);$$

this was done by Marshall and Olkin (1960a).

#### 4.6 Bivariate distribution: convex polygonal region

As an illustration of what can be achieved by transforming variables, we show how to obtain  $L$  when  $T$  is the plane convex polygonal region, symmetrical about the origin, formed by the intersection of  $m$  strips derived by rotation of  $|x| \leq d_i$  through angles  $\alpha_i$  ( $i = 1, \dots, m$ ). We suppose that means are zero and that variances and covariances are given. It will be noted that if in Section 4.2 we take  $n = 2$  then we have the special case of the present problem in which  $m = 2$ ,  $\alpha_1 = 0$ , and  $\alpha_2 = \frac{1}{2}\pi$ .

Hence we know that if  $T$  is  $|x_1| < 1$ ,  $|x_2| < 1$  and

$$E(x'x) = \begin{pmatrix} Ex_1^2 & Ex_1x_2 \\ Ex_1x_2 & Ex_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = M,$$

then  $L$  is  $\max [0, 1 - \frac{1}{2}(\sigma_1^2 + \sigma_2^2 + \sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4\rho^2\sigma_1^2\sigma_2^2})]$ , and that we find this value for  $L$  by taking for  $g(x_1, x_2)$  the expression  $1 + (x_1, x_2)A(x_1, x_2)'$  where  $A^{-1}$  is

$$\begin{pmatrix} -1 & \beta \\ \beta & -1 \end{pmatrix}$$

and  $\beta = \{ -(\sigma_1^2 + \sigma_2^2) + \sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4\rho^2\sigma_1^2\sigma_2^2} \} / 2\rho\sigma_1\sigma_2$ .

In this case  $g = 0$  is an ellipse which touches the lines  $x_1 = \pm 1$ ,  $x_2 = \pm 1$ .

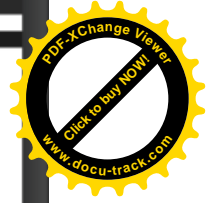
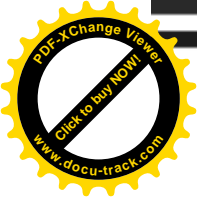
If we now have the region

$$|c_{11}x_1 + c_{12}x_2| \leq d_1, \quad |c_{21}x_1 + c_{22}x_2| \leq d_2,$$

then we first put

$$(y_1, y_2) = (x_1, x_2) \begin{pmatrix} c_{11}/d_1 & c_{21}/d_2 \\ c_{12}/d_1 & c_{22}/d_2 \end{pmatrix}$$

or  $y = xC$ , and  $g$  becomes  $1 + xCAC'x'$  with



$$A = \begin{pmatrix} -1 & \beta \\ \beta & -1 \end{pmatrix}^{-1}$$

and  $\beta$  derived from the values, not in  $M$ , but in  $C'MC$ .

If the ellipse  $g = 0$  lies inside all the other strips then the value of  $L$  obtained is the correct one for the polygon; if this does not happen then we must consider ellipses which lie inside the hexagons formed by the intersection of three strips. Again we can take a particular case and obtain the general one by transformation, and we therefore suppose that the hexagon is defined by

$$|x_1| \leq 1, |x_2| \leq 1, |x_1 \cos \alpha + x_2 \sin \alpha| \leq d.$$

By changing the sign of  $x_2$  if necessary we may suppose that  $0 < \alpha < \frac{1}{2}\pi$ , and in order that the region be strictly a hexagon, we need  $(d^2 - 1)^2 < \sin^2 2\alpha$ . Let  $(d^2 - 1) = \gamma \sin 2\alpha$ .

The ellipse  $Q(x_1, x_2) = x_1^2 + x_2^2 - 2\gamma x_1 x_2 - 1 + \gamma^2 = 0$  is inscribed to the hexagon and touches the sides at  $A(1, \gamma)$ ,  $B(\gamma, 1)$  and  $C((\gamma \sin \alpha + \cos \alpha)/d, (\gamma \cos \alpha + \sin \alpha)/d)$  respectively and also at  $-A$ ,  $-B$  and  $-C$  (the reflections of  $A$ ,  $B$  and  $C$  in the origin). Outside the hexagon we have  $Q > 0$  and everywhere  $Q \geq \gamma^2 - 1$ , so that if  $P$  is the probability of  $(x_1, x_2)$  being inside the hexagon we have

$$E(Q) = \sigma_1^2 + \sigma_2^2 - 2\gamma\rho\sigma_1\sigma_2 - 1 + \gamma^2 \geq P(\gamma^2 - 1)$$

or  $P \geq P_0 = 1 - (\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2\gamma)/(1 - \gamma^2)$ .

We now show that this bound (if  $P_0 \geq 0$ ) is sharp.

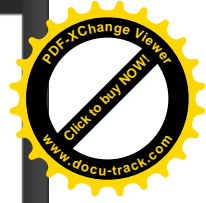
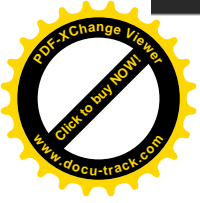
Consider the discrete distribution with probabilities  $\frac{1}{2}p_1$  at  $A$  and  $-A$ ,  $\frac{1}{2}p_2$  at  $B$  and  $-B$ ,  $\frac{1}{2}p_3$  at  $C$  and  $-C$ , and  $1 - p_1 - p_2 - p_3$  at the origin, where

$$p_1 = \frac{\sigma_1^2(\gamma \cos \alpha + \sin \alpha) + \sigma_2^2(\gamma^2 \sin \alpha + \gamma \cos \alpha) - \rho\sigma_1\sigma_2(\gamma^2 \cos \alpha + 2\gamma \sin \alpha + \cos \alpha)}{\sin \alpha (1 - \gamma^2)^2}$$

$$p_2 = \frac{\sigma_1^2(\gamma^2 \cos \alpha + \gamma \sin \alpha) + \sigma_2^2(\gamma \sin \alpha + \cos \alpha) - \rho\sigma_1\sigma_2(\gamma^2 \sin \alpha + 2\gamma \cos \alpha + \sin \alpha)}{\cos \alpha (1 - \gamma^2)^2}$$

$$p_3 = \frac{(-\gamma\sigma_1^2 - \gamma\sigma_2^2 + \rho\sigma_1\sigma_2(1 + \gamma^2))(1 + \gamma \sin 2\alpha)}{\sin \alpha \cos \alpha (1 - \gamma^2)^2}.$$

E



Then all moments are correct and  $1 - p_1 - p_2 - p_3 = P_0$ .

Now since we are considering a hexagon and not merely a parallelogram, the ellipses obtained above for the parallelograms formed by any two of the strips must each intersect the sides of the third strip which determines the hexagon; in particular the ellipse  $-1 = (x_1, x_2) A (x_1, x_2)'$  intersects the line  $d = x_1 \cos \alpha + x_2 \sin \alpha$  if  $(\beta + \gamma) \sin \alpha \cos \alpha \leq 0$ , so that

$$p_3 = p \sigma_1 \sigma_2 (\gamma + \beta)(\gamma + \beta^{-1}) / \sin \alpha \cos \alpha (1 - \gamma^2)^2 \geq 0;$$

similarly  $p_1 > 0$ ,  $p_2 > 0$  and the distribution we have considered is an actual one.

If  $P_0 < 0$  we can decrease  $\sigma_1$  and  $\sigma_2$  until  $P_0 = 0$  and construct a discrete distribution with probabilities on the boundary of the hexagon and none inside. Then restoring  $\sigma_1$  and  $\sigma_2$  to their original values has the effect of moving these probabilities away from the origin and so giving zero probability in  $T$ . Thus we have proved that

$$L = \max(0, P_0).$$

For a general polygon we have to consider first all possible parallelograms and then, if these fail, all possible hexagons, selecting the least bound given.

This problem was originally discussed by Marshall and Olkin (1960c); they showed in addition that if the widths of the strips are equal the only hexagons to be considered are those formed by adjacent strips, in the sense that no other strip defining the polygon is obtained by a rotation of the first strip through an angle less than those defining the hexagon.

#### 4.7 A general theorem (first- and second-order moments given)

In this section we give a theorem, due to Marshall and Olkin (1960b), which reduces the problem of finding bounds on probability in certain problems to the minimization of a quadratic form under given conditions. As in the problem studied in Section 4.2 the solution may not be explicit, but we give one application where it is.

We suppose that  $T$  is a closed convex set and  $T^*$  the union of  $T$  and  $-T$ , where  $-T$  is the image of  $T$  in the origin (means are chosen to be zero). Let the variance-covariance matrix  $E(x'x)$  be denoted by  $M$ .

If  $A$  is the set of all vectors  $a$  such that  $ax' \geq 1$  for all  $x$  in  $T$ , then for  $T$  we have

$$U = \inf_{a \text{ in } A} \frac{aMa'}{1 + aMa'} \quad (4.7.1)$$

while for  $T^*$

$$U = \inf_{a \text{ in } A} aMa'. \quad (4.7.2)$$

To prove (4.7.1) we consider  $g(x) = (ax' + aMa')^2 / (1 + aMa')^2$ . We have  $g(x) \geq 1$  for  $x$  in  $T$  and so obtain

$$aMa' / (1 + aMa') = E(g(x)) \geq \Pr(x \text{ in } T). \quad (4.7.3)$$

Now let  $q = q(a) = aMa'$ ,  $Q = q/(1 + q)$ ,  $w = aM/q$ . Using Cauchy's inequality<sup>(\*)</sup>, we have  $xMx'wM^{-1}w' \geq (xw')^2$ , whence  $xMx' \geq q(xw')^2 \geq Q(xw')^2 = Qxw'wx'$ . Hence  $x(M - Qw'w)x' \geq 0$ , and equality can obtain only if  $Q = q$  or  $xw' = 0$ , which means that  $q = 0$  or  $xw' = 0$ ; therefore equality obtains only when  $xMx' = 0$ . Since  $M$  is the matrix of a positive definite form so is  $M - Qw'w$ . Hence there is a non-singular matrix  $N$  such that  $M - Qw'w = N'N$ . Since  $xMx'$  is positive definite we shall have  $\inf aMa'$  taken at some finite point, say  $a_0$ . We now show that  $w_0 = w(a_0)$  belongs to  $T$ .

If this is not so then, since  $T$  is convex and closed, there is a hyperplane separating  $w_0$  and  $T$ ; i.e. there is a vector  $v$  and a constant  $\alpha$  such that  $vw' < \alpha$  but  $vx' \geq \alpha$  if  $x$  is in  $T$ . Since  $a_0w_0' = 1$ , we have  $(v + (1 - \alpha)a_0)w_0' < 1$  and also  $(v + (1 - \alpha)a_0)x' \geq 1$  for  $x$  in  $T$ , and we may therefore suppose that  $\alpha = 1$  by taking  $v + (1 - \alpha)a_0$  in place of  $v$ . Now  $vw_0' = vMa'/q$  and so  $vMa'_0 < a_0Ma'_0$ , i.e.  $a_0Mv' < a_0Ma'_0$ . Hence for a sufficiently small positive  $\epsilon$ , we shall have  $\epsilon(vMv' - 2a_0Mv' + a_0Ma'_0) < 2(a_0Ma'_0 - a_0Mv')$  i.e.

$$uMu' < a_0Ma'_0 \quad (4.7.4)$$

(\*)For this form of the inequality see, e.g., Beckenbach and Bellman (1961), p. 69.

where  $u = \epsilon v + (1 - \epsilon) a_0$ . However, if  $x$  belongs to  $T$  we have  $ux' \geq \epsilon + (1 - \epsilon) = 1$  and so (4.7.4) contradicts the definition of  $a_0$ . Hence  $w_0$  is in  $T$ .

We now show that the bound obtained for probability in (4.7.3) is sharp, so that we have the actual value of  $U$ .  $Q, w$ , etc. are all to be evaluated at  $a_0$  and we shall drop suffixes.

Let  $D = \text{diag}(p_1^{\frac{1}{2}}, \dots, p_n^{\frac{1}{2}})$  (i.e. the square matrix with  $p_1^{\frac{1}{2}}, \dots, p_n^{\frac{1}{2}}$  on the diagonal and zero elsewhere) where  $eD = -QwN^{-1}$  and  $e = (1, \dots, 1)$ , and let  $C = D^{-1}N$ . We have  $\sum p_i = eD^2 e' = Q^2 wN^{-1} N^{-1} w' = Q^2 w(M - Qw'w)^{-1} w'$ . Now since  $qw' = Ma'$  and  $wa' = 1$ , we have  $w(M - Qw'w)^{-1}(M - Qw'w)a' = 1$ , whence  $w(M - Qw'w)^{-1} w' = 1/(q - Q) = (1 - Q)/Q^2$  and so  $\sum p_i = 1 - Q$ . Let  $C^{(i)}$  be the  $i$ th row of  $C$ , and consider the discrete distribution with probabilities  $p_i$  at  $C^{(i)}$  and  $Q$  at  $w$ . This distribution has mean  $wQ + eD^2 C = wQ + eDN = wQ - Qw = 0$  and variance-covariance matrix  $Qw'w + C'D^2 C = Qw'w + N'N = M$  and so satisfies all the given conditions. Since the probability at  $w$  (which belongs to  $T$ ) is  $Q$ , we have  $U \geq Q$ , whereas from (4.7.3) we have  $U \leq Q$ . Hence  $U = Q$ .

We note that since  $N$  is not unique the distribution found to give the value for  $U$  is not unique.

For the region  $T^*$  the argument is similar; we take  $g(x) = (ax')^2$  (and so do not use the fact that  $E(x) = 0$ ) and we find that  $M - qw'w$  is positive semi-definite, so that now  $N$  is singular. We choose positive  $p_1, \dots, p_n$  to satisfy  $p_1 + \dots + p_n = 1 - q$  and put  $C = D^{-1}N$  where  $D = \text{diag}(p_1^{\frac{1}{2}}, \dots, p_n^{\frac{1}{2}})$ . Finally we split the probabilities  $p_1, \dots, p_n, q$  equally between  $C^{(i)}$  and  $-C^{(i)}$  or  $w$  and  $-w$ .

As an application we take  $T$  to be the region in which every component of  $x$  is not less than 1; thus  $T$  is the region in which all components have the same sign and modulus not less than 1. We look for the minimum of  $aMa'$  subject to  $a \geq 0, ae' \geq 1$ ; we must have  $a \geq 0$  since if  $a_1$ , say, were negative we could find a point  $(X, 1, \dots, 1)$  in  $T$  which gave  $ax' < 1$  for sufficiently large  $X$ . If the minimum occurs at an  $a$  with non-vanishing components  $b$  then we can take  $be' = 1$  (here  $e$  is a row of 1's, as many as there are elements

in  $b$ ); differentiation of  $bM_s b' + \lambda(b e' - 1)$  (where  $M_s$  is the principal submatrix of  $M$  corresponding to the columns in  $b$ ) gives  $2M_s b' + \lambda e' = 0$ . Since  $b e' = 1$  we have  $2bM_s b' + \lambda = 0$  and so  $b' = -\frac{1}{2} \lambda M_s^{-1} e' = M_s^{-1} e' (bM_s b')$ .

$$\text{Hence } bM_s b' = (bM_s b')^2 eM_s^{-1} M_s M_s^{-1} e'$$

$$\text{i.e. } bM_s b' = 1/(eM_s^{-1} e')$$

$$\text{and so } b' = M_s^{-1} e'/(eM_s^{-1} e').$$

Hence  $\min aMa'$  is given by  $1/\max eM_s^{-1} e'$  where  $M_s^{-1} e' > 0$ . This maximum always exists since a possible  $M_s$  is given by any diagonal term of  $M$ ; it will be noticed that finding the minimum involves only the consideration of a finite number of submatrices, and not, as in the case of the problem discussed in Section 4.2, the solution of sets of equations.

Marshall and Olkin also apply the general theorem to the case when  $T$  is  $x_1 \dots, x_n \geq 1, x > 0$ , but then the solution is not explicit as was the one which we have just obtained.

#### 4.8 Elliptical region: independent variables

In the above results we obtain equality only when the distribution is concentrated on a quadric or at its centre; if we stipulate that the variables should be independently distributed then we exclude these cases.

Suppose that  $T$  is  $x_1^2/s_1^2 + x_2^2/s_2^2 = 1$  and the variances of  $x_1, x_2$  are  $\sigma_1^2, \sigma_2^2$  respectively where  $\sigma_1^2/s_1^2 \leq \sigma_2^2/s_2^2$ . The method which we have used before to obtain inequalities breaks down here since when we write down the expectation of  $x_1 x_2$  we use only the fact that  $x_1$  and  $x_2$  are uncorrelated and not the stronger condition that they are independent. Birnbaum, Raymond and Zuckerman (1947) solved the problem by approximating to a distribution by a discrete distribution with probability at many points and then reducing the number of points to at most four. We state the result in terms of non-negative variables (which may be taken to be  $x_1^2/s_1^2$  and  $x_2^2/s_2^2$  for the present problem) as follows. Let  $x$  and  $y$  be non-negative with  $E(x) = \lambda, E(y) = \mu$  and  $\lambda \leq \mu$ .

Then  $\Pr(x + y \geq t) \leq M(t)$  where

$$\begin{aligned} M(t) &= 1 \quad \text{if } t \leq \lambda + \mu, \\ &= \mu/(t - \lambda) \quad \text{if } \lambda + \mu \leq t \leq \frac{1}{2}(\lambda + 2\mu + \sqrt{(\lambda^2 + 4\mu^2)}), \\ &= (\lambda + \mu)/t - \lambda\mu/t^2 \quad \text{if } \frac{1}{2}(\lambda + 2\mu + \sqrt{(\lambda^2 + 4\mu^2)}) \leq t. \end{aligned}$$

The inequality is realized with equality for certain discrete distributions.

We first suppose that  $x_1, x_2, x_3$  are three values of  $x$  (now assumed to have a discrete distribution) at which the probabilities are  $p_1, p_2, p_3$  respectively. We shall replace these probabilities by  $p'_1, p'_2, p'_3$  such that  $p'_1 p'_2 p'_3 = 0$ ,

$$p'_1 + p'_2 + p'_3 = p_1 + p_2 + p_3 \quad (4.8.1)$$

$$x_1 p'_1 + x_2 p'_2 + x_3 p'_3 = x_1 p_1 + x_2 p_2 + x_3 p_3, \quad (4.8.2)$$

and, for the new distribution,  $P = \Pr(x \geq u)$  is not less than its value with the original distribution.

From (4.8.1) and (4.8.2) we have

$$p'_1 = p_1 + \frac{(x_3 - x_2)}{(x_3 - x_1)} (p_2 - p'_2), \quad p'_3 = p_3 + \frac{(x_2 - x_1)}{(x_3 - x_1)} (p_2 - p'_2).$$

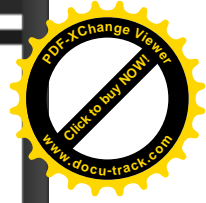
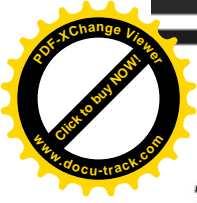
If  $u$  does not lie in the interval  $x_1 \leq u \leq x_3$  then we have not changed  $P$ ; if  $x_1 \leq u \leq x_2$  then we can increase  $P$  by decreasing  $p_1$  and this we do by taking

$$p'_2 - p_2 = \min \left( p_1 \frac{x_3 - x_1}{x_3 - x_2}, p_3 \frac{x_3 - x_1}{x_2 - x_1} \right),$$

while if  $x_2 \leq u \leq x_3$  we increase  $p_3$  by taking  $p'_2 = 0$ .

Now suppose that, starting from discrete distributions which approximate as closely as we wish to the actual distributions of  $x$  and  $y$ , we have successively reduced the number of points with non-zero probability until we have, for  $x$ ,  $p$  at  $\alpha$  and  $1 - p$  at  $\beta$  ( $\alpha < \beta$ ) while, for  $y$ , we have  $q$  at  $\gamma$  and  $1 - q$  at  $\delta$  ( $\gamma < \delta$ ), where

$$\lambda = p\alpha + (1 - p)\beta, \quad \mu = q\gamma + (1 - q)\delta. \quad (4.8.3)$$



Then, in the combined distribution, we have probabilities  $pq$  at  $A(\alpha, \gamma)$ ,  $p(1-q)$  at  $B(\alpha, \delta)$ ,  $(1-p)q$  at  $C(\beta, \gamma)$  and  $(1-p)(1-q)$  at  $D(\beta, \delta)$ .

If  $t \leq \lambda + \mu$  then the example  $p = q = 0$  shows that  $M(t) \geq 1$  and so  $M(t) = 1$ ; suppose in what follows that  $t > \lambda + \mu$ . If  $t < \beta$  we can, without altering  $E(x)$ , replace the distribution of  $x$  by  $p$  at  $\alpha' = \alpha + (1-p)(\beta - t)/p$  and  $1-p$  at  $t$ ; note that  $t - \alpha' = (t - \lambda)/p > 0$  so that  $\Pr(x \geq t)$  is unaltered. Hence we may suppose that  $\beta \leq t$  and similarly  $\delta \leq t$ .

If, at  $A$ ,  $x + y \leq t$  while at  $B, C, D$ ,  $x + y \geq t$  then

$$\begin{aligned} P &= 1 - pq = 1 - (\beta - \lambda)(\delta - \mu)/(\beta - \alpha)(\delta - \gamma) \quad (\text{from (4.8.3)}) \\ &\leq 1 - (\beta - \lambda)(\delta - \mu)/(\beta + \delta - t)^2 \end{aligned}$$

(since  $\alpha + \delta \geq t$ ,  $\beta + \gamma \geq t$ ). The right-hand side, regarded as a function of  $\beta$ , has a single minimum and so its greatest value is

$$\max(1 - (t - \mu - \lambda)/(\delta - \mu), 1 - (t - \lambda)(\delta - \mu)/\delta^2)$$

since  $t - \mu \leq t - \gamma \leq \beta \leq t$ .

Now  $1 - (t - \mu - \lambda)/(\delta - \mu) \leq \lambda/(t - \mu) \leq \mu/(t - \lambda)$  since  $\delta \leq t$ , and  $1 - (t - \lambda)(\delta - \mu)/\delta^2$ , which, as a function of  $\delta$ , has a single minimum, is not more than

$$\max(1 - (t - \lambda - \mu)/(t - \lambda), 1 - (t - \lambda)(t - \mu)/t^2)$$

(since  $t - \lambda \leq t - \alpha \leq \delta \leq t$ ),

$$= \max(\mu/(t - \lambda), (\mu + \lambda)/t - \mu\lambda/t^2).$$

If  $A$  and  $B$  lie on the same side of  $x + y = t$  as the origin and  $C, D$  lie on the opposite side, then

$$\begin{aligned} P &= 1 - p = 1 - (\beta - \lambda)/(\beta - \alpha) \\ &= (\lambda - \alpha)/(\beta - \alpha) \leq (\lambda - \alpha)/(t - \mu - \alpha) \\ &\leq \lambda/(t - \mu) \leq \mu/(t - \lambda). \end{aligned}$$

If  $A$  and  $C$  lie on the same side of  $x + y = t$  as the origin, and  $B$  and  $D$  lie on the opposite side, then

$$\begin{aligned} P &= 1 - q = 1 - (\delta - \mu)/(\delta - \gamma) \\ &= (\mu - \gamma)/(\delta - \gamma) \leq (\mu - \gamma)/(t - \lambda - \gamma) \leq \mu/(t - \lambda). \end{aligned}$$

If  $A, B$  and  $C$  lie on the same side of  $x + y = t$  as the origin and  $D$  lies on the opposite side, then

$$P = (1 - p)(1 - q) = (\lambda - \alpha)(\mu - \gamma)/(\beta - \alpha)(\delta - \gamma)$$

and if  $\beta + \delta = t'$  then

$$P \leq \lambda\mu/\beta\delta = \lambda\mu/\beta(t' - \beta) \leq \max(\mu/(t' - \lambda), \lambda/(t' - \mu)),$$

since  $\lambda \leq \beta \leq t' - \mu$ , i.e.

$$P \leq \mu/(t' - \lambda) \leq \mu/(t - \lambda).$$

Hence for  $\lambda + \mu < t$  we have

$$P \leq \max(\mu/(t - \lambda), (\lambda + \mu)/t - \lambda\mu/t^2)$$

which is  $\mu/(t - \lambda)$  if  $t \leq \frac{1}{2}(\lambda + 2\mu + \sqrt{(\lambda^2 + 4\mu^2)})$  and is  $(\lambda + \mu)/t - \lambda\mu/t^2$  if  $t \geq \frac{1}{2}(\lambda + 2\mu + \sqrt{(\lambda^2 + 4\mu^2)})$ .

These bounds can be attained; e.g.  $\alpha = \lambda, \beta = t, \gamma = 0, \delta = t - \lambda$  give  $P = \mu/(t - \lambda)$ , while  $\alpha = \gamma = 0, \beta = \delta = t$  give  $P = (\lambda + \mu)/t - \lambda\mu/t^2$ .

Without the hypothesis of independence the best result we can obtain is  $P \leq (\lambda + \mu)/t$ .

Although it would be possible in principle to use the above method for the sum of more than two variables it is clear that there would be many more cases to consider and the working would be long. The principle of reduction can be employed when more than one expectation is given, but working with more than three values at a time and so ending with more than two with non-zero probabilities; this was done by Hoeffding (1955). The particular case when the  $x$ 's have the same distribution was discussed by these means by Hoeffding and Shrikhande (1955).

#### 4.9 A monotonicity condition: ellipsoidal region

In the previous sections of this chapter we have used only moments of the distributions. A restriction on the shape of the distribution was introduced by Leser (1942).

We take  $T$  to be the region  $\sum (x_i/\lambda_i \sigma_i)^2 \leq n$ .

Let  $n \lambda_0^{-2} = \sum \lambda_i^{-2}$ ,  $n \sigma_0^{-2} = \sum \sigma_i^{-2}$ ,  $R^2 = (\lambda_0^2/n) \sum (x_i/\lambda_i \sigma_i)^2$  (so that  $T$  is  $R \leq \lambda_0$ ).

We let  $A(R_0)$  be the mean value of the p.d.f. on the ellipsoid  $R = R_0$  and suppose that  $A(R)$  is a non-increasing function of  $R$  for  $R \leq K$ . We thus have a condition analogous to that imposed by Narumi (see Exercise 12).

If the surface content of the ellipsoid  $(\lambda_0^2/n) \sum (x_i/\lambda_i \sigma_i)^2 = R^2$  is  $CR^{n-1}$  then we have

$$1 = \int_0^\infty CR^{n-1} A(R) dR, \quad 1 = \int_0^\infty CR^{n+1} A(R) dR,$$

$$P = \int_0^{\lambda_0} CR^{n-1} A(R) dR.$$

Put  $u = CA(\lambda_0)/n$  and  $K_0 = \{\lambda_0^n + (1-P)/u\}^{1/n}$ .

If  $K \geq K_0$ , i.e.

$$P \geq 1 - u(-\lambda_0^n + K^n), \quad (4.9.1)$$

then since  $K \geq \lambda_0$  we have

$$1 = \int_0^{\lambda_0} CR^{n+1} A(R) dR + \int_{\lambda_0}^\infty CR^{n+1} A(R) dR = I_1 + I_2, \text{ say.}$$

Further, we have  $I_1 \geq \int_0^{\lambda_0} CA(\lambda_0) R^{n+1} dR$ .

Also

$$\int_{\lambda_0}^\infty CA(R) R^{n-1} dR = 1 - P = \int_{\lambda_0}^{K_0} CA(\lambda_0) R^{n-1} dR$$

$$\text{or } \int_{K_0}^\infty CA(R) R^{n-1} dR = \int_{\lambda_0}^{K_0} CR^{n-1} (A(\lambda_0) - A(R)) dR. \quad (4.9.2)$$

The integrands on both sides are non-negative, and the values of  $R$  in the left-hand integrand are larger than the values of  $R$  in the right-hand integrand. Hence, if we multiply both integrands by  $R^2$  we obtain

$$\int_{K_0}^\infty CA(R) R^{n+1} dR \geq \int_{\lambda_0}^{K_0} CR^{n+1} (A(\lambda_0) - A(R)) dR$$

$$\text{or } I_2 \geq CA(\lambda_0) \int_{\lambda_0}^{K_0} R^{n+1} dR$$

and so  $1 \geq \int_0^{K_0} CA(\lambda_0) R^{n+1} dR = CA(\lambda_0) K_0^{n+2}/(n+2),$

$$\text{whence } P \geq 1 + \lambda_0^n u - u((n+2)/nu)^{n/(n+2)}. \quad (4.9.3)$$

Also from  $K \geq \lambda_0$  we have

$$P \geq \int_0^{\lambda_0} CR^{n-1} A(\lambda_0) dR = \lambda_0^n u. \quad (4.9.4)$$

If  $\lambda_0 < K < K_0$  then

$$P \leq 1 + u(\lambda_0^n - K^n); \quad (4.9.5)$$

(4.9.4) still holds, but the integrand on the right-hand side of (4.9.2) is no longer necessarily non-negative.

We now write

$$\begin{aligned} \int_{\lambda_0}^{\infty} CA(R) R^{n-1} dR &= n(1-P) \left( \int_{\lambda_0}^K R^{n-1} dR \right) / (K^n - \lambda_0^n) \\ &= CA(\lambda_0)(K_0^n - \lambda_0^n) \left( \int_{\lambda_0}^K R^{n-1} dR \right) / (K^n - \lambda_0^n) \\ &= \int_{\lambda_0}^K R^{n-1} dR CA(\lambda_0) + \left( \int_{\lambda_0}^K R^{n-1} dR \right) \\ &\quad (K_0^n - K^n) CA(\lambda_0) / (K^n - \lambda_0^n). \end{aligned}$$

$$\begin{aligned} \text{Hence } \int_K^{\infty} CA(R) R^{n-1} dR &= \int_{\lambda_0}^K CR^{n-1} (A(\lambda_0) - A(R)) dR \\ &\quad + \left( \int_{\lambda_0}^K R^{n-1} dR \right) CA(\lambda_0) (K_0^n - K^n) / (K^n - \lambda_0^n). \end{aligned}$$

We multiply the integrands by  $R^2, R^2, K^2$  respectively and obtain

$$\begin{aligned} I_2 &\geq \int_{\lambda_0}^K CA(\lambda_0) R^{n+1} dR + \\ &\quad \left( \int_{\lambda_0}^K R^{n-1} dR \right) CA(\lambda_0) (K_0^n - K^n) K^2 / (K^n - \lambda_0^n) \end{aligned}$$

whence

$$\begin{aligned} 1 &\geq \int_0^K CA(\lambda_0) R^{n+1} dR + \\ &\quad \left( \int_{\lambda_0}^K R^{n-1} dR \right) CA(\lambda_0) (K_0^n - K^n) K^2 / (K^n - \lambda_0^n), \end{aligned}$$

i.e.

$$1 \geq K^{n+2} CA(\lambda_0)/(n+2) + K^2 CA(\lambda_0)(K_0^n - K^n)/n,$$

whence

$$P \geq 1 - K^{-2} + (\lambda_0^n - 2K^n/(n+2))u. \quad (4.9.6)$$

Finally

$$I_1 \geq 0, I_2 \geq \lambda_0^2 \int_{\lambda_0}^{\infty} CR^{n-1} A(R) dR = \lambda_0^2 (1 - P),$$

so that  $P \geq 1 - \lambda_0^{-2}$ , but this bound is improved on by the inequalities above whenever they are applicable.

For  $\lambda_0 \leq K$  we now have to find the minimum, as  $u$  varies, of the lower bound for  $P$ , using the sets of inequalities (a) (4.9.1), (4.9.3) and (4.9.4) or (b) (4.9.4), (4.9.5) and (4.9.6).

If  $K \leq 1$ , (b) with  $u = 0$  give  $P = 0$ , and (a) can give no less.

If  $\lambda_0^n \geq 2K^n/(n+2)$  then the least bound from (b) is  $1 - K^{-2}$  for  $u = 0$ , and since the bound in (4.9.3) has minimum

$$1 - (2/(n+2))^{2/n} \lambda_0^{-2} \text{ for } u = \left(\frac{n+2}{n}\right) \left(\frac{2}{n+2}\right)^{(n+2)/n} \lambda_0^{-(n+2)} \text{ and}$$

this is not less than  $1 - K^{-2}$ , we have  $P \geq 1 - K^{-2}$ .

Now suppose that  $\lambda_0^n \leq 2K^n/(n+2)$ .

If  $1 \leq K \leq \sqrt{\{(n+2)/n\}}$  then the lower bound from (b) is

$$\lambda_0^n (n+2) (1 - K^{-2})/2K^n \text{ for } u = (n+2) (1 - K^{-2})/2K^n.$$

If  $\sqrt{\{(n+2)/n\}} \leq K$  then this bound is inadmissible on account of (4.9.5), and we have instead

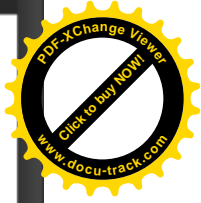
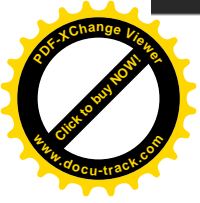
$$1 + (n+2)(\lambda_0^n - K^n)/nK^{n+2} \text{ for } u = (n+2)/n K^{n+2}.$$

Also if  $1 \leq K \leq \sqrt{\{(n+2)/n\}}$  the lower bound from (a) is  $(\lambda_0/K)^n$  for  $u = K^{-n}$ , while if  $\sqrt{\{(n+2)/n\}} \leq K$  the lower bound from (a) is  $1 - \{2/(n+2)\}^{2/n} \lambda_0^{-2}$  if

$$\lambda_0 \geq \{2/(n+2)\}^{1/n} \{(n+2)/n\}^{1/2}$$

and  $\lambda_0^n \{n/(n+2)\}^{n/2} \text{ for } u = \{n/(n+2)\}^{1/2}$  if

$$\lambda_0 \leq \{2/(n+2)\}^{1/n} \{(n+2)/n\}^{1/2}.$$



Combining these results, we have as the lower bound for  $P$ :

$$0 \quad \text{if } \lambda_0 \leq 1, K \leq 1.$$

$$1 - \lambda_0^{-2} \quad \text{if } K \leq \lambda_0.$$

$$1 - K^{-2} \quad \text{if } 2K^n/(n+2) \leq \lambda_0^n \leq K^n.$$

$$\lambda_0^n (n+2)(1 - K^{-2})/2K^n \quad \text{if } 1 \leq K \leq \sqrt{\{(n+2)/n\}}$$

and

$$\lambda_0 \leq 2^{1/n} K (n+2)^{-1/n}.$$

$$n^{n/2} \lambda_0^n (n+2)^{-n/2} \quad \text{if } \sqrt{\{(n+2)/n\}} \leq K$$

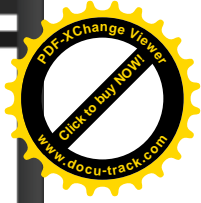
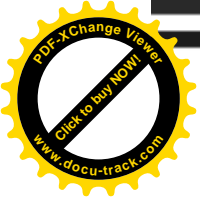
and

$$\lambda_0 \leq \left(\frac{2}{n+2}\right)^{1/n} \left(\frac{n+2}{n}\right)^{\frac{1}{2}}.$$

$$1 - \left(\frac{2}{n+2}\right)^{2/n} \lambda_0^{-2} \quad \text{if } \sqrt{\left(\frac{n+2}{n}\right)} \leq K$$

and

$$\left(\frac{2}{n+2}\right)^{1/n} \left(\frac{n+2}{n}\right)^{\frac{1}{2}} \leq \lambda_0 \leq \left(\frac{2}{n+2}\right)^{1/n} K.$$



## CHAPTER V

### SUMS OF VARIABLES

#### 5.1 Introduction

In this chapter we consider not single variables but the sums of a number of variables, not necessarily all having the same distribution. In a sense we are dealing with a special case of multivariate distributions, since the joint p.d.f. of a sample  $x_1, \dots, x_n$  can be regarded as a multivariate p.d.f. for the point  $(x_1, \dots, x_n)$ , and restrictions on the sum  $x_1 + \dots + x_n$  (or on partial sums) define regions in the sample space. The sum, however, has a particular interest owing to the Central Limit Theorem which states that under certain conditions the distribution of a sum tends to normality as the number of variables summed tends to infinity, and for that reason results obtained in this connection have been collected in a separate chapter (see Section 5.4 for a more detailed reference to the Central Limit Theorem). As in previous chapters we concentrate on methods which yield definite numerical bounds and usually ignore results which contain undetermined constants. We assume the  $x$ 's to be independently distributed unless the contrary is stated.

#### 5.2 Population and sample variances given

The inequality which follows, due to Guttman (1948b), is unusual in that both the population variance and the sample variance are used in it.

If the average of  $x_1, \dots, x_n$  is  $\bar{x}$  and the maximum likelihood (biased) estimate of the sample variance is  $s^2 = \sum (x_i - \bar{x})^2 / n$  then we have

$$\begin{aligned} E((\bar{x} - \mu'_1)^2) &= \mu_2/n, & E((\bar{x} - \mu'_1)^4) &= (\mu_4/n^3) + 3(n-1)\mu_2^2/n^3, \\ E(s^2) &= (n-1)\mu_2/n, & E(s^4) &= (n-1)^2\mu_4/n^3 + \\ & & & (n^2 - 2n + 3)(n-1)\mu_2^2/n^3, \end{aligned}$$

and

$$E((\bar{x} - \mu'_1)^2 s^2) = (n-1)\mu_4/n^3 + (n-3)(n-1)\mu_2^2/n^3.$$

Hence if we put  $u = (\bar{x} - \mu'_1)^2 - s^2/(n-1) - c\mu_2$  then we have  $E(u^2) = \mu_2^2(c^2 + 2/n(n-1))$ . Hence, from Markov's inequality (see Exercise 3) we have

$$\Pr \left\{ |u| \leq \lambda \mu_2 \sqrt{\left( \frac{2}{n(n-1)} + c^2 \right)} \right\} \geq 1 - \lambda^{-2},$$

i.e.

$$\Pr \left\{ (\bar{x} - \mu'_1)^2 \leq \frac{s^2}{n-1} + c\mu_2 + \lambda \mu_2 \sqrt{\left( \frac{2}{n(n-1)} + c^2 \right)} \right\} \geq 1 - \lambda^{-2}. \quad (5.2.1)$$

(We have now possibly increased the probability by including the range

$0 \leq (\bar{x} - \mu'_1)^2 \leq s^2/(n-1) + c\mu_2 - \lambda \mu_2 \sqrt{(c^2 + 2/n(n-1))}$  if this exists.)

To minimize the bound for  $(\bar{x} - \mu'_1)^2$  in (5.2.1) we take  $c < 0$  and  $c^2 = 2/n(n-1)(\lambda^2 - 1)$  to give

$$\Pr \left\{ (\bar{x} - \mu'_1)^2 \leq \frac{s^2}{n-1} + \mu_2 \sqrt{\left( \frac{2(\lambda^2 - 1)}{n(n-1)} \right)} \right\} \geq 1 - \lambda^{-2}. \quad (5.2.2)$$

A straightforward application of Markov's inequality to  $(\bar{x} - \mu'_1)$  would give

$$\Pr \{ |\bar{x} - \mu'_1| \leq \lambda \sqrt{(\mu_2/n)} \} \geq 1 - \lambda^{-2} \quad (5.2.3)$$

so that we have replaced the bound  $\lambda^2 \mu_2/n$  by

$$\frac{s^2}{n-1} + \mu_2 \sqrt{\left( \frac{2(\lambda^2 - 1)}{n(n-1)} \right)}.$$

If  $\mu_2 = 1$ ,  $s = 1$ ,  $n = 20$ , then (5.2.3) gives

$$\Pr \{ (\bar{x} - \mu'_1)^2 \leq 1 \} \geq \frac{19}{20} = .95,$$

while (5.2.2) gives

$$\Pr \{ (\bar{x} - \mu'_1)^2 \leq 1 \} \geq \frac{3240}{3259} = .9942 \dots$$

### 3 Symmetrical unimodal bounded distribution

In the case when  $x_i$  has a symmetrical unimodal bounded distribution (not necessarily the same for each  $i$ ) we can obtain an inequality by showing that the rectangular distribution is the most extreme case. We suppose that  $f(x_i)$  is zero for  $|x_i| > a$ , and let  $y_k$  be the sum of  $k$  variables distributed in the rectangular distribution with mean zero and range 2. We now prove, by induction on  $n$ , that

$$\Pr(|\sum_{i=1}^n x_i| > ba) \leq \Pr(|y_n| > b) \quad \text{for all } b \geq 0.$$

Let  $F(t)$ ,  $G(t)$  be respectively the distribution functions of  $x_i/a$  and of the rectangular distribution with mean zero and range 2; since the p.d.f. of  $x$  is non-increasing in the interval  $0 \leq x_i \leq a$  the graph of  $F(t)$  is concave downwards for  $0 \leq t \leq 1$ , and since it passes through the points  $(0, \frac{1}{2})$  and  $(1, 1)$  on the graph of  $G(t)$  (which is a straight line between these points) we have

$$F(t) \geq G(t) \quad \text{for } 0 \leq t \leq 1.$$

Hence  $\Pr(x_i > ba) = 1 - F(b) \leq 1 - G(b) = \Pr(y > b)$ .

The reasoning for negative values of the variables is similar, and this establishes the truth of the proposition for  $n = 1$ .

Now let  $F_n$  be the distribution function of  $(x_1 + \dots + x_n)/a$ ,  $F$  that of  $x_{n+1}/a$ ,  $G_n$  that of  $y_n$ , and  $G$  that of  $(y_{n+1} - y_n)$ .

For  $b > 0$  we now have

$$\begin{aligned} & - \Pr\left(\sum_{i=1}^{n+1} x_i > ba\right) + \Pr(y_{n+1} > b) \\ &= \int_{-\infty}^{\infty} \{F_n(b-s) dF(s) - G_n(b-s) dG(s)\} \\ &= \int_{-\infty}^{\infty} F_n(b-s) (dF(s) - dG(s)) + \\ & \quad + \int_{-\infty}^{\infty} (F_n(b-s) - G_n(b-s)) dG(s) \\ &= \int_{-\infty}^{\infty} (F(b-s) - G(b-s)) dF_n(s) + \\ & \quad + \int_{-\infty}^{\infty} (F_n(b-s) - G_n(b-s)) dG(s), \quad (5.3.1) \end{aligned}$$

(on integrating by parts in the first integral and then replacing  $s$  by  $b - s$ ).

We can now write the first integral as

$$\int_0^\infty (F(s) - G(s)) dF_n(b - s) + \int_0^\infty (F(-s) - G(-s)) dF_n(b + s), \quad (5.3.2)$$

by dividing the interval of integration into ranges  $-\infty$  to  $b$  and  $b$  to  $\infty$  and replacing  $s$  by  $b - s$  or  $b + s$  respectively. Now because of the symmetry of the distributions of  $x_{n+1}$  and of  $(y_{n+1} - y_n)$  we have  $F(-s) = 1 - F(s)$  and  $G(-s) = 1 - G(s)$ , so that the expression in (5.3.2) becomes

$$\int_0^\infty (F(s) - G(s)) (dF_n(b - s) - dF_n(b + s)).$$

Now  $F(s) - G(s) \geq 0$  for  $0 \leq s$ , while

$$dF_n(b - s) - dF_n(b + s) \geq 0 \quad \text{if } 0 \leq s \leq b,$$

since the derivative of  $F_n(s)$  is non-increasing for positive  $s$ , and if  $b \leq s$  then

$$dF_n(b - s) - dF_n(b + s) = dF_n(s - b) - dF_n(s + b)$$

by the symmetry of  $dF_n(s)$ , so that again

$$dF_n(b - s) - dF_n(b + s) \geq 0.$$

Hence the first integral in (5.3.1) is non-negative and so, similarly, is the second. This completes the proof by induction, which was given, in a more general form, by Birnbaum (1948). The need for the monotonicity condition is shown in Exercise 18.

Now  $\Pr(|y_n| > b)$  is

$$\frac{2}{n!} \sum_{\frac{n}{2}(b+1) < k \leq n} (-1)^k {}^nC_k \left\{ \frac{n}{2}(b+1) - k \right\}^n \quad (5.3.3)$$

(see, e.g., Kendall and Stuart (1958, 1963), p. 257) and so

$$\Pr\left(\left|\sum_1^n x_i\right| > ba\right)$$

is not greater than the expression in (5.3.3).

## 5.4 Comparison with normal distribution

The sum of a number of random variables tends, as the number tends to infinity, to have the normal distribution under certain conditions (see, e.g., Loève (1955), Chapter VI). It is thus possible to approximate to the probability that the sum should fall in a certain interval by means of the known probability that a normal variable should do so. A bound for the difference between the distribution functions seems to have been given first by Liapounoff in a form involving an unknown constant; later writers have given numerical values for the constant and so enabled a definite numerical result to be obtained by use of the theorem. Since the analysis which leads to the results is too long to be reproduced here we merely give results and references.

Cramér (1923) and (1928) proved that if the  $x_i$  are independent with  $E(x_i) = 0$ ,  $E(x_i^2) = \sigma_i^2$ ,  $E(|x_i|^3) = \tau_i^3$  ( $i = 1, \dots, n$ ) and  $s_n^2 = \sigma_1^2 + \dots + \sigma_n^2$ ,  $t_n = \tau_1^3 + \dots + \tau_n^3$ , then the distribution function of  $x = \sum x_i/s_n$  satisfies

$$\left| F_n(x) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt \right| < \frac{3 t_n \log n}{s_n^3} \quad \text{for } n > 1. \quad (5.4.1)$$

Using this result Offord (1945) showed that the probability of  $x_1 + \dots + x_n$  lying in an interval of length  $2\lambda$  is not greater than

$$\frac{6 \log n}{k^3 \sqrt{n}} \left( \log n + \frac{k\lambda}{\min \sigma_i} \right) \quad (5.4.2)$$

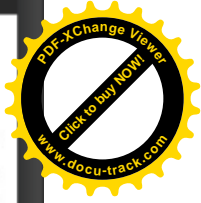
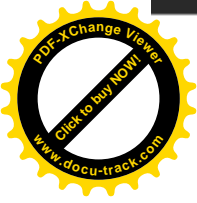
where  $\sigma_i/\tau_i \geq 2 K^\dagger$ .

Moreover  $n, k, \min \sigma_i$  can be replaced in (5.4.2) by the corresponding quantities for a subsequence of at least two terms from  $x_1, \dots, x_n$ . (But the inequality still refers to the sum of all the  $x$ 's.)

Bergström (1949) replaced the factor  $3 \log n$  in (5.4.1) by  $4.8$  and also gave a result for the case when the  $x$ 's are not independent.

Instead of using the sum of the third moments, Berry (1941) replaced the right-hand side of (5.4.1) by

$$\frac{1.88}{s_n} \max_i \frac{\tau_i^3}{\sigma_i^2}.$$



This result is better than Bergström's when the distributions of the  $x_i$  are identical, but not necessarily so if they are different. Berry's proof contains errors (see Hsu (1945), p. 3), and the correctness of the constant 1.88 has been disputed. The value 2.031 has been given by Takano (1950). In the case when the distributions of the  $x_i$  are the same, Ikeda (1959) has obtained better values of the constant by requiring that  $t_n/s_n^3$  shall not be too small.

To end the section we show that the removal of the  $\log n$  term in (5.4.1) has produced a right-hand side of the correct order; further improvement must lie in the direction of improving the constant. Let the  $x_i$  have the binomial distribution with probabilities  $\frac{1}{2}$  at  $x = \pm 1$ , so that  $\sigma_i = \tau_i = 1$ ,  $s_n = \sqrt{n}$ , and  $t_n = n$ . For even  $n$  the probability that  $\sum x_i = 0$  is  ${}^nC_{\frac{1}{2}n} 2^{-n}$  which is asymptotically  $(2\pi n)^{-\frac{1}{2}}$  by using Stirling's formula for the factorials. Consequently at points near the origin on either side

$$\left| F_n(x) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt \right|$$

must be about  $\frac{1}{2} (2\pi n)^{-\frac{1}{2}}$ , so that the term  $n^{-\frac{1}{2}}$  is of the correct order and the constants 1.88, etc. cannot be less than  $\frac{1}{2}(2\pi)^{-\frac{1}{2}} = .199 \dots$

### 5.5 Restrictions on all moments

A result under the condition that the rate of growth of moments be not too large was given by Bernstein (1924). Suppose that for the distribution of each  $x_i$  there exists a constant  $H$  such that  $|\mu_r| \leq \frac{1}{2} H r^{-2} (r!) \mu_2$  for  $2 \leq r$ . Note that if the range of  $x_i$  is finite so that  $|x_i| \leq M$ , then we have

$$|\mu_r| \leq \int_{-\infty}^{\infty} M r^{-2} x^2 f(x) dx = M r^{-2} \mu_2$$

and

$$\frac{2|\mu_r|}{r! \mu_2} \left(\frac{3}{m}\right)^{r-2} \leq \frac{2.3^{r-2}}{r!} \leq 1,$$

so that we may take  $H = M/3$ .

$$\text{Now } E(e^{\theta x}) = E\left(1 + \theta x + \frac{\theta^2 x^2}{2!} + \dots\right)$$

and

$$\begin{aligned} E(e^{\theta x}) &\leq 1 + \frac{\theta^2 \mu_2}{2!} + \sum_3^{\infty} \frac{\theta^r H^{r-2} r! \mu_2}{2(r!)} \\ &= 1 + \frac{\theta^2 \mu_2}{2!} \cdot \frac{1}{1 - \theta H}, \end{aligned}$$

if there exists  $c$  such that  $H|\theta| \leq c < 1$ .

Hence

$$E(e^{\theta x}) \leq e^{\theta^2 \mu_2 / 2(1-c)} \quad \text{and} \quad E(e^{\theta \sum x_i}) \leq e^{\theta^2 S_n^2 / 2(1-c)}$$

(with  $S_n^2$  defined as in Section 5.4), so that

$$\Pr \{e^{\theta \sum x_i} > e^{\lambda^2 + \theta^2 S_n^2 / 2(1-c)}\} \leq e^{-\lambda^2},$$

i.e.

$$\Pr \left\{ \sum x_i > \frac{\lambda^2}{\theta} + \frac{\theta S_n^2}{2(1-c)} \right\} \leq e^{-\lambda^2}.$$

To minimize the bound we take  $\theta^2 = 2(1-c)\lambda^2/S_n^2$  to give

$$\Pr \{ \sum x_i > S_n \lambda / \sqrt{2(1-c)} \} \leq e^{-\lambda^2}$$

and for the result to be valid we need  $2(1-c)\lambda^2/S_n^2 \leq c^2/H^2$ . If, for example, we choose  $c = \frac{1}{2}$ , then we have

$$\Pr \{ \sum x_i > \lambda s_n \} \leq e^{-\lambda^2} \quad \text{for } 0 < \lambda < s_n/2H.$$

This theorem has been modified by Bernstein (1937), using the idea of expectation of  $x_k$  relative to  $x_1, \dots, x_{k-1}$  to remove the restriction that the  $x_i$  be independent. Modifications due to Craig (1933) are to take moments over an arbitrarily large, but finite, interval  $-b \leq x \leq b$  or to work with cumulants instead of moments.

## 5.6 An inequality for partial sums

Instead of working with the sum  $x_1 + \dots + x_n$ , as in the earlier sections of this chapter, we may consider all the partial sums  $x_1, x_1 + x_2, \dots$  and seek a bound for the probability that these lie in certain intervals. Such a bound was first given by Kolmogoroff (1928) (see Kolmogoroff (1929) for corrections to parts of this paper),

but we give here the generalized result due to Hájek and Rényi (1955), which is itself a special case of a still more general theorem of Birnbaum and Marshall (1961).

We suppose that  $x_1, x_2, \dots$  is a sequence of mutually independent random variables with zero means and finite variances  $\sigma_k^2 = E(x_k^2)$ .  $c_1, c_2, \dots$  is a non-increasing sequence of positive numbers.

If

$$z = \sum_{k=n}^{m-1} (x_1 + \dots + x_k)^2 (c_k^2 - c_{k+1}^2) + c_m^2 (x_1 + \dots + x_m)^2$$

then

$$\begin{aligned} E(z) &= \sum_{k=n}^{m-1} (c_k^2 - c_{k+1}^2) (\sigma_1^2 + \dots + \sigma_k^2) + c_m^2 (\sigma_1^2 + \dots + \sigma_m^2) \\ &= c_n^2 \sum_{k=1}^n \sigma_k^2 + \sum_{k=n+1}^m c_k^2 \sigma_k^2. \end{aligned}$$

For a value  $r$  such that  $n \leq r \leq m$  let  $E_r$  be the event

$$|x_1 + \dots + x_s| < \epsilon/c_s \quad (n \leq s < r), \quad |x_1 + \dots + x_r| \geq \epsilon/c_r.$$

We now consider expectations of various quantities on the hypothesis that  $E_r$  has occurred.

We have, for  $r < i \leq m$ ,  $E(x_i | E_r) = 0$ , so that

$$\begin{aligned} E((x_1 + \dots + x_k)^2 | E_r) &= E((x_1 + \dots + x_r)^2 + \\ &\quad + 2(x_1 + \dots + x_r)(x_{r+1} + \dots + x_k) + (x_{r+1} + \dots + x_k)^2 | E_r) \\ &\geq E((x_1 + \dots + x_r)^2 | E_r) \geq \epsilon^2/c_r^2 \quad (r \leq k \leq m). \end{aligned}$$

Hence

$$E(z | E_r) \geq \sum_{k=r}^{m-1} \epsilon^2 (c_k^2 - c_{k+1}^2)/c_r^2 + c_m^2 \epsilon^2/c_r^2 = \epsilon^2$$

and

$$E(z) \geq \sum_{r=n}^m E(z | E_r) P(E_r) \geq \epsilon^2 \sum_{r=n}^m P(E_r).$$

Hence

$$\begin{aligned} \sum_n^m P(E_r) &= \Pr \left\{ \max_{n \leq k \leq m} c_k |x_1 + \dots + x_k| \geq \epsilon \right\} \\ &\leq (c_n^2 \sum_1^n \sigma_k^2 + \sum_{n+1}^m c_k^2 \sigma_k^2) / \epsilon^2. \end{aligned} \quad (5.5.1)$$

Kolmogorov proved the special case in which  $n = 1$  and all the  $c_k$  are 1. If we take  $c_k = 1/k$  and  $\sigma_k = \sigma$  then we have

$$\Pr \left\{ \max_{n \leq k \leq m} |x_1 + \dots + x_k| \geq \epsilon k \right\} \leq \frac{\sigma^2}{\epsilon^2} \left( \frac{1}{n} + \sum_{n+1}^m \frac{1}{k^2} \right) < \frac{2\sigma^2}{n\epsilon^2}.$$

Hence  $|x_1 + \dots + x_m|/m$  converges to zero ( $= E(x_k)$ ) with probability one; this is a form of the strong law of large numbers (see Loève (1955)).

Marshall (1960) showed that for the one-sided inequality we have

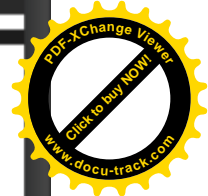
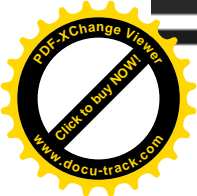
$$\Pr \left\{ \max_{1 \leq i \leq n} (x_1 + \dots + x_i) \geq \epsilon \right\} \leq \frac{(\sigma_1^2 + \dots + \sigma_n^2)}{(\epsilon^2 + \sigma_1^2 + \dots + \sigma_n^2)}.$$

This was proved by putting

$$x = \left( \epsilon \sum_1^n x_i + \sigma_1^2 + \dots + \sigma_n^2 \right)^2 / (\epsilon^2 + \sigma_1^2 + \dots + \sigma_n^2),$$

taking  $E_r$  as  $x_1 + \dots + x_s < \epsilon$  ( $1 \leq s < r$ ),  $x_1 + \dots + x_r \geq \epsilon$  and proceeding as before.

Marshall also discussed the question of introducing multipliers in the way Hájek and Rényi did, but showed that even for  $n = 2$  the result was complicated.

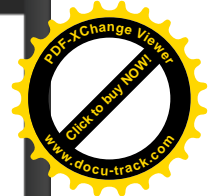
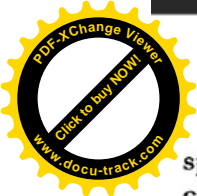


## CHAPTER VI

### APPLICATIONS

In this chapter we consider what use may be made of the material in the previous chapters, not only by the statistician but by the pure mathematician. For the latter the chief interest of any survey of work must be as a source of ideas for further work — in filling in gaps in the existing theory, in extending existing ideas to cover new or more general situations, or in embedding the whole theory in some wider set of ideas. The scope for further work increases as we go through the monograph; in Chapter II (apart from the last section) there is effectively a complete solution to the problem of finding bounds for probability, and all that is needed is to devise ways of reducing the amount of computation involved (though this is, of course, a far from trivial matter). All this is, however, done under the assumption that bounds exist; obviously, if moments arise from an actual distribution there is at least one value for the probability in a given set, but if we start from an arbitrarily chosen set of numbers it is less easy to decide whether they are realizable with a distribution; this applies with more force in Chapter III where fewer of the restrictive conditions on data such as (2.2.2) are known. Mallows (1956) suggests (in his Theorem III, which is really a conjecture) that the conditions are realizable if his method leads to bounds, and the same seems likely to be true for the methods in this monograph (Exercise 1 gives slight support to this view.) The existence of generality does not obtain in Sections 2.10 and 3.7, where we deal only with very specialized problems, and there is considerable scope for further work.

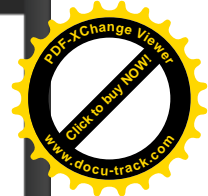
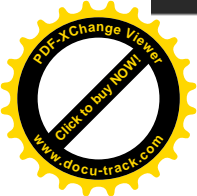
In Chapter IV the situation changes radically; here, when a general method applies it leads to computational problems (as in Section 4.2) of a different order of difficulty, and this using only second-order moments and a simple type of region. Only by



specializing the problems still further in some of the later sections can we obtain explicit solutions. Each of the sections of this chapter is to be regarded only as a sample of the type of complexity likely to be met with when more general methods are developed. The work of Whittle (1958a) is particularly interesting in taking what amounts to the limiting case of a multivariate distribution as its dimension becomes infinite and introducing quite different techniques and conditions.

In Chapter V again we have little generality but only a collection of special results, although by the exclusion of results expressed in phrases such as "... all sufficiently large ..." in favour of results giving definite numerical values we have ignored a large body of intricate work.

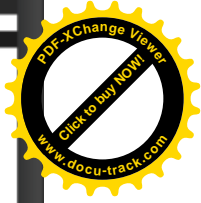
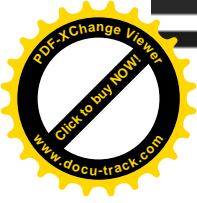
If the pure mathematician finds the subject full of gaps the statistician may find it all too full of material; his requirement is for a formula which will, as speedily as possible, produce a useful result from his data. The aim of the section headings has been to indicate where results of a particular type may be found (the exercises associated with a section should be consulted at the same time), but for quick reference an excellent résumé of the simpler results has been provided by Savage (1961). Savage also gives three examples of the application of inequalities; the first relates to the heights of soldiers, which are known to lie between two bounds, and the probability that the mean of a sample differ by more than an inch from the population mean is required. This problem can be solved by using the fact that the variable is bounded to give an estimate of its variance and then using Tchebychef's inequality or, more accurately, Bernstein's inequality. The second example relates to the magnitude of cumulative sums; variance is supposed to be known from past experience, and Kolmogorov's inequality is used. The third example relates to correlated variables; the coefficient of correlation is supposed to be known, and Berge's inequality is used. An application of a different type was suggested by Barton (in the discussion following the paper by Mallows (1956)) who noted that the distributions of test functions under hypotheses alternative to the null hypothesis might not have many numerical parameters known.



but might reasonably be supposed to be “smooth” in the sense that the p.d.f. and its derivatives are not very different from what we have with the (known) distribution under the null hypothesis. Using these ideas, he showed that quite close estimates can be obtained of the number of trials necessary to give conclusions with a given degree of confidence.

It should be noted that in all but the first of the four examples we assume some knowledge of the distribution — either actual values of parameters or bounds on the p.d.f. or its derivatives. In the first we avoid this by obtaining for a bounded distribution crude estimates of the parameters. Some such assumption of knowledge is inevitable, since all inequalities are expressed partly or wholly in terms of population parameters and not the sample estimates of them. Consequently we have the paradoxical situation that we can use the inequalities most effectively when we know so much about the population that use of the inequalities is unnecessary. At other times we work under the tacit assumption that if our estimates are not too far from the population values then our conclusions will not be too far out; effectively, when we state a result such as “the probability is at most . . .” we are suppressing the supplementary statement “and the first statement is true with probability . . .”.

A further use of the theory is to study the extent to which knowledge of a distribution is relevant. Thus in Section 2.6 it was found that the effect of moments above the fourth on the bounds  $L$  and  $U$  was small compared with the effect of the first four moments, and hence there would be little point in trying to estimate the higher moments.



## EXERCISES

The chapter or section to which each exercise is most closely related is shown in brackets after the number of the exercise. The reader is recommended to construct and solve for himself examples such as those worked in Sections 2.4 and 3.5 and to verify the numerical results on pages 25, 26, 28 and 29.

1. (2.3) If  $\mu'_2 = 1$  and  $T$  is  $0 \leq x$  show that  $\psi$  has no infimum if  $|\mu'_1| > 1$ .

2. (2.6) If  $F_1(x)$  and  $F_2(x)$  are the distribution functions of distributions with the same first  $2n$  moments prove that

$$|F_1(k) - F_2(k)| \leq \left| \begin{array}{ccc} 1 & \dots & \mu_n(a) \\ & \dots & \\ \mu_n(a) & \dots & \mu_{2n}(a) \end{array} \right| \left| \begin{array}{ccc} \mu_2(k) & \dots & \mu_{n+1}(k) \\ & \dots & \\ \mu_{n+1}(k) & \dots & \mu_{2n}(k) \end{array} \right|$$

for any value of  $a$ . (Khamis (1954).)

3. (2.7) If  $T$  is  $x \geq k > 0$  and  $\nu_1$  is given, prove that  $U = \nu_1/k$ . (This is Markov's inequality.)

4. (2.9)  $h(x)$  is a positive function of  $x$  with minimum value  $H$  and is increasing for  $x \geq k \geq 0$ .

$$m_h = \int_0^\infty \{f(x) + f(-x)\} h(x) dx,$$

and  $T$  is  $|x| \geq k$ . Prove that  $m_h \geq H$  and sketch the region in which  $(a_0, a_1)$  lies if  $a_0 + a_1 h(x) \geq \chi_T(x)$ . Hence prove that  $U = \min(1, (m_h - H)/(h(k) - H))$ , whence  $U \leq m_h/h(k)$ . (The last result was given by Cantelli (1910).)

5. (2.9) If  $T$  is  $0 \leq x \leq k$  and  $\nu_n, \nu_{2n}$  are given, prove that for  $0 \leq k^n \leq \nu_n$ ,  $L = 0$ ,  $U = (\nu_{2n} - \nu_n^2)/(\nu_{2n} - \nu_n^2 + (k^n - \nu_n)^2)$ ; for  $\nu_n \leq k^n \leq \nu_{2n}/\nu_n$ ,  $L = 1 - (\nu_n/k^n)$ ,  $U = 1$ ; for  $\nu_{2n} \leq k^n \leq \nu_n$ ,  $L = (k^n - \nu_n)^2/(\nu_{2n} - \nu_n^2 + (k^n - \nu_n)^2)$ ,  $U = 1$ . (Cantelli (1928).)

6. (II)  $n$  mutually independent trials have  $s$  mutually exclusive results;  $p_i$  is the probability of the  $i$ th result, and  $q_i$  the observed relative frequency. Show, using Markov's inequality (Exercise 3), that

$$\Pr \left\{ \sum (p_i - q_i)^2 < \lambda^2 \right\} > 1 - (s - 1)/sn\lambda^2.$$

If  $n'$  and  $n''$  trials give  $q'_i, q''_i$  for the relative frequencies, show that

$$\Pr \left\{ \sum (q'_i - q''_i)^2 < \lambda^2 \right\} > 1 - (s - 1)(n' + n'')/sn'n''\lambda^2.$$

(Romanovski (1940).)

7. (II) If  $x$  is a random variable in  $(0, 2\pi)$  and  $E(\sin x) = \alpha$ ,  $E(\cos x) = \beta$ , then

$$\Pr(2\theta < x < 2\phi) \geq \frac{\alpha \sin(\theta + \phi) + \beta \cos(\theta + \phi) - \cos(\phi - \theta)}{1 - \cos(\phi - \theta)}$$

$$\Pr(2\theta \leq x \leq 2\phi) \geq \frac{\alpha \sin(\theta + \phi) + \beta \cos(\theta + \phi) + 1}{1 + \cos(\phi - \theta)}$$

where  $0 \leq \theta \leq \phi \leq \pi$ . (Marshall and Olkin (1961).)

8. (II) If  $\delta$  is the mean deviation about the mean  $\mu$  and  $T$  is

$$\mu - t_1 \delta \leq x \leq \mu + t_2 \delta, \text{ then } P(T) \geq 1 - \frac{1}{2} \left( \frac{1}{t_1} + \frac{1}{t_2} \right).$$

Show that if  $\delta$  is less than the standard deviation (by Schwarz's inequality it cannot be greater), then for values of  $t_1$  and  $t_2$  near unity the inequality using mean deviation is sharper than the one given in Section 2.5. (Glasser (1961).)

9. (3.2) Show that if  $f(x)$  has a single maximum at  $x = 0$  and  $v_n$  is given, then

$$L = k/((n+1) v_n)^{1/n} \quad \text{if } k \leq n((n+1) v_n)^{1/n}/(n+1);$$

$$L = 1 - v_n n^n/k^n (n+1)^n \quad \text{if } n((n+1) v_n)^{1/n}/(n+1) \leq k.$$

(The history of these inequalities is given in Fréchet (1950).)

10. (3.2)  $x$  is a non-negative variable,  $f(x)$  has a single maximum at  $c$ ,  $v_1$  is given, and  $T$  is  $0 \leq x \leq k$ . By consideration of

$$\int_0^\infty x(x-c)f'(x)dx \quad \text{prove that } c \leq 2v_1.$$

$$\text{Let } a = (2k - c - 2\sqrt{k(k-c)})/c^2.$$

$$\text{If } v_1 < c < k \text{ show that } L = 1 + (c - 2v_1)a.$$

$$\text{If } c < v_1, c < k \text{ and } b = (k-c)/4(c-v_1)^2 < a, \text{ show that}$$

$$L = 2b(c-v_1) + \sqrt{4b(k-c)}.$$

$$\text{If } k < c \leq 2v_1 - k \text{ show that } L = 0.$$

$$\text{If } k < c, 2v_1 - k < c \text{ show that } L = (k + c - 2v_1)/c.$$

11. (3.2)  $\mu'_1 = 0$ ,  $\mu'_2 = \frac{1}{2}k^2$ ,  $f(x)$  has a single maximum at  $x = k$ , and  $T$  is  $|x| \leq k$ . Prove that  $L = 64/81$ ,  $U = 1$ .

12. (III)  $x$  is a non-negative variable,

$$m^n = \int_0^\infty x^n f(x) dx,$$

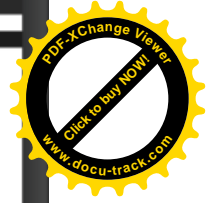
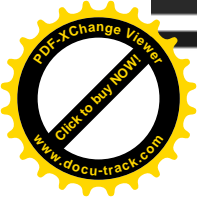
and  $z$  is defined as a function of  $y$  by

$$y = \int_0^{zm} f(x) dx = F(zm).$$

Prove that  $z$  is a non-decreasing function of  $y$  and that

$$1 = \int_0^1 z^n dy.$$

If  $f(x)$  is non-decreasing for  $0 \leq x \leq bm$  ( $1 < b < (n+1)^{1/n}$ ) prove that the graph of  $z$  against  $y$  is concave downwards for  $0 < z < b$  and hence that



$$\begin{aligned}
0 &\leq F(km) \leq k/b \quad \text{for } 0 \leq k \leq b_1, \\
k/b - (b-k)(n+1-b^n)/b(b^n-k^n) \\
&\leq F(km) \leq k/b \quad \text{for } b_1 \leq k \leq b, \\
(k^n-1)(n+1)/((n+1)k^n-b^n) \\
&\leq F(km) \leq 1 \quad \text{for } b \leq k,
\end{aligned}$$

where  $b_1$  is the positive root of  $k(b_1^n - k^n) = (b_1 - k)(n + 1 - b_1^n)$ .

(Narumi (1923); note that the restriction on  $f(x)$  is over a range from the origin and not on the tail of the distribution for large  $x$ , as in the work of von Mises (1938).)

13. (III) Replace the condition in Exercise 12 by  $f(x)$  non-increasing and show that

$$z \leq nb^{n+1}y/(n+1)(b^n-1) \quad \text{for } 0 \leq y \leq 1 - b^{-n}$$

and hence that

$$\begin{aligned}
F(km) &\geq (n+1)(b^n-1)k/nb^{n+1} \quad \text{for } 0 \leq k \leq nb/(n+1), \\
&\geq 1 - b^{-n} \quad \text{for } nb/(n+1) \leq k \leq b \\
&\geq 1 - k^{-n} \quad \text{for } b \leq k.
\end{aligned}$$

(Narumi (1923).)

14. (III)  $x$  is a non-negative variable and  $f(x)$  is non-increasing for  $v_1 \leq x$ . Show that

$$\int_{x=0}^{\infty} (x - v_1)^2 dF(x) > \int_{F=F_0}^1 \frac{(k - v_1)^2 (F - F_1)^2}{(F(k) - F_1)^2} dF,$$

where the tangent to the graph of  $F(x)$  at  $x = k$  meets  $x = v_1$  at  $(v_1, F_1)$ . Deduce that

$$F(t\sigma) > 1 - 4(\sigma^2 - v_1^2)/9(k - v_1)^2.$$

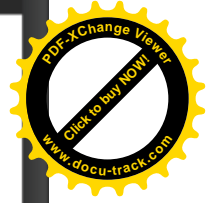
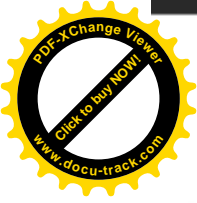
(Peck 1933).)

15. (4.2) A matrix equivalent to

$$\begin{pmatrix} -1 & t & t \\ t & -1 & t \\ t & t & -1 \end{pmatrix}$$

was proposed for  $A$  by Lal (1955); show that this gives a best possible result only if all correlation coefficients are equal to

$$(t^2 - 2t)/(1 + 2t^2).$$



16. (4.8)  $x_1, \dots, x_{2n}$  are independent with means zero and variances  $\sigma^2$ . Prove that

$$\begin{aligned}\Pr \left\{ \sum_1^{2n} x_i^2 \geq \lambda n \sigma^2 \right\} &\leq 1 \quad \text{for } \lambda \leq 2 \\ &\leq 1/(\lambda - 1) \quad \text{for } 2 \leq \lambda \leq \frac{1}{2}(3 + \sqrt{5}) \\ &\leq \frac{2}{\lambda} \left( 1 - \frac{1}{2\lambda} \right) \quad \text{for } \frac{1}{2}(3 + \sqrt{5}) \leq \lambda.\end{aligned}$$

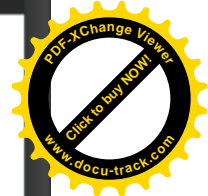
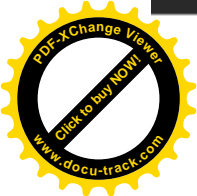
(Birnbaum, Raymond and Zuckerman (1947).)

17. (4.8) If non-negative variables  $x_1$  and  $x_2$  have the same distribution with mean  $\mu$  prove that

$$\begin{aligned}\Pr \{(x_1 + x_2) > t\mu\} &\leq (2/t)^2 \quad \text{for } 2 \leq t \leq 5/2 \\ &\leq 2/t - 1/t^2 \quad \text{for } 5/2 \leq t.\end{aligned}$$

Show that with its stronger hypothesis this gives an improvement over Birnbaum, Raymond and Zuckerman's inequality (Exercise 16) for  $5/2 \leq t \leq \frac{1}{2}(3 + \sqrt{5})$ .

18. (5.3) If  $x_1$  and  $x_2$  have probability  $\frac{1}{2}$  at  $x = a$  and  $x = -a$ , show that the proposition proved in (5.3) is false. (Birnbaum (1948).)



## BIBLIOGRAPHY

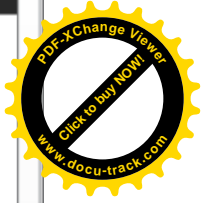
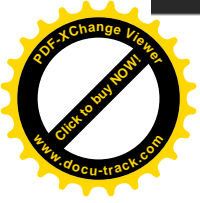
Previous surveys of the subject have been given by Fréchet (1950) (Chapter IV) and the present author (1955). An extensive bibliography was given by Savage (1953) who later (1961) gave some further references. The list which follows contains only work referred to in the text or the exercises and is not intended to be exhaustive. Work which is considered to have been superseded or to be of minor importance is not listed. A few of the references are given from quotation in other work.

Numbers in square brackets refer to the pages on which items are mentioned.

- BECKENBACH, E. F. and BELLMAN, R. (1961). *Inequalities*. Springer. [67]
- BERGE, P. O. (1937). A note on a form of Tchebycheff's theorem for two variables. *Biometrika*, **29**, 405-406. [61]
- BERGSTRÖM, H. (1949). On the central limit theorem in the case of not identically distributed random variables. *Skandinavisk Aktuarietidskrift*, **32**, 37-62. [81]
- BERNSTEIN, S. (1924). Sur une modification de l'inégalité de Tchebichef. *Ann. Sc. Inst. Sav. Ukraine, Sect. Math. I.* (Russian, French summary.) [82]
- BERNSTEIN, S. (1937). Sur quelques modifications de l'inégalité de Tchebycheff. *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS*, **17**, 279-282. [83]
- BERRY, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.*, **49**, 122-136. [81]
- BIRNBAUM, Z. W. (1948). On random variables with comparable peakedness. *Ann. Math. Statist.*, **19**, 76-81. [80, 93]
- BIRNBAUM, Z. W. and MARSHALL, A. W. (1961). Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. *Ann. Math. Statist.*, **32**, 687-703. [84]
- BIRNBAUM, Z. W., RAYMOND, J. and ZUCKERMAN, H. S. (1947). A generalization of Tshebyshev's inequality to two dimensions. *Ann. Math. Statist.*, **18**, 70-79. [69, 93]
- CANTELLI, F. P. (1910). Intorno ad un teorema fondamentale della teoria del rischio. *Bollettino dell' Associazione degli Attuari*, Milan, 1-23. [89]
- CANTELLI, F. P. (1928). Sui confini della probabilità. *Att. Cong. Intern. Mat.*, **6**, Bologna, 47-60. [90]
- CRAIG, C. C. (1933). On the Tchebycheff inequality of Bernstein. *Ann. Math. Statist.*, **4**, 94-102. [83]
- CRAMÉR, H. (1923). Das Gesetz von Gauss und die Theorie des Risikos. *Skandinavisk Aktuarietidskrift*, **6**, 209-237. [81]
- CRAMÉR, H. (1928). On the composition of elementary errors. *Skandinavisk Aktuarietidskrift*, **11**, 13-74 and 141-180. [81]
- EGGLESTON, H. G. (1958). *Convexity*. Cambridge Univ. Press. [10]

- FRÉCHET, M. (1950). *Généralités sur les Probabilités: Éléments Aléatoires* (2nd edn). Gauthier-Villars, Paris. [39, 91, 94]
- GALE, D. (1960). *The theory of linear economic models*. McGraw-Hill. [17]
- GLASSER, G. J. (1961). Tchebycheff-type inequalities in terms of the mean deviation. *Sankhyā*, Series A, **23**, 397–400. [90]
- GODWIN, H. J. (1955). On generalizations of Tchebychef's inequality. *J. Amer. Statist. Soc.*, **50**, 923–945. [94]
- GUTTMAN, L. (1948a). An inequality for kurtosis. *Ann. Math. Statist.*, **19**, 277–278. [18]
- GUTTMAN, L. (1948b). A distribution-free confidence interval for the mean. *Ann. Math. Statist.*, **19**, 410–413. [77]
- HÁJEK, J. and RÉNYI, A. (1955). Generalization of an inequality of Kolmogorov. *Acta Math. Sci. Hung.*, **6**, 281–284. [84]
- HOEFFDING, W. (1955). The extrema of the expected value of a function of independent random variables. *Ann. Math. Statist.*, **26**, 268–275. [72]
- HOEFFDING, W. and SHRIKHANDE, S. S. (1955). Bounds for the distribution function of a sum of independent, identically distributed random variables. *Ann. Math. Statist.*, **26**, 439–449. [72]
- HSU, P. L. (1945). The approximate distribution of the mean and variance of a sample of independent variables. *Ann. Math. Statist.*, **16**, 1–29. [82]
- IKEDA, S. (1959). A note on the normal approximation to the sum of independent random variables. *Ann. Inst. Statist. Math. Tokyo*, **11**, 121–130. [82]
- ISII, K. (1959a). On a method for generalizations of Tchebychef's inequality. *Ann. Inst. Statist. Math. Tokyo*, **10**, 65–88. [18]
- ISII, K. (1959b). Bounds for probability for non-negative random variables. *Ann. Inst. Statist. Math. Tokyo*, **11**, 89–99. [27]
- KENDALL, M. G. and STUART, A. (1958, 2nd edn. 1963). *The Advanced Theory of Statistics*, Vol. I. Charles Griffin & Co., London. [15, 33, 80]
- KHAMIS, S. H. (1954). On the reduced moment problem. *Ann. Math. Statist.*, **25**, 113–122. [89]
- KINGMAN, J. F. C. (1963). On inequalities of the Tchebychev type. *Proc. Cambridge Phil. Soc.*, **59**, 135–146. [18]
- KOLMOGOROV, A. N. (1928). Über die Summen durch den Zufall bestimmter unabhängiger Grössen. *Math. Ann.*, **99**, 309–319. [83]
- KOLMOGOROV, A. N. (1929). Bemerkungen zu meiner Arbeit "Über die Summen zufälligen Grössen". *Math. Ann.*, **102**, 484–488. [83]
- LAL, D. N. (1955). A note on a form of Tchebychef's inequality for two or more variables. *Sankhyā*, **15**, 317–320. [61, 92]
- LESER, C. E. V. (1942). Inequalities for multivariate frequency distributions. *Biometrika*, **32**, 284–293. [72]
- LOÈVE, M. (1955). *Probability Theory*. Van Nostrand. [62, 81, 85]
- MALLOWS, C. L. (1956). Generalizations of Tchebycheff's inequalities. *J. Roy. Statist. Soc. (B)*, **18**, 139–168. Discussion on 168–176. [45, 48, 86, 87]
- MARSHALL, A. W. (1960). A one-sided analogue of Kolmogorov's inequality. *Ann. Math. Statist.*, **31**, 483–487. [85]
- MARSHALL, A. W. and OLKIN, I. (1960a). A one-sided inequality of the Chebyshev type. *Ann. Math. Statist.*, **31**, 488–491. [61, 64]

- MARSHALL, A. W. and OLKIN, I. (1960b). Multivariate Chebyshev inequalities. *Ann. Math. Statist.*, **31**, 1001-1014. [18, 66, 69]
- MARSHALL, A. W. and OLKIN, I. (1960c). A bivariate Chebyshev inequality for symmetric convex polygons. *Contributions to probability and statistics: Essays in Honor of Harold Hotelling*. Stanford Univ. Press. 299-308. [66]
- MARSHALL, A. W. and OLKIN, I. (1961). Game theoretic proof that Chebyshev inequalities are sharp. *Pacific J. of Math.*, **11**, 1421-1429. [2, 90]
- MIRSKY, L. (1955). *Introduction to Linear Algebra*. Oxford Univ. Press. [14, 23]
- NARUMI, S. (1923). On further inequalities with possible applications to problems in the theory of probability. *Biometrika*, **15**, 245-253. [73, 92]
- OFFORD, A. C. (1945). An inequality for sums of independent random variables. *Proc. London Math. Soc.* (2), **48**, 467-477. [81]
- OLKIN, I. and PRATT, J. W. (1958). A multivariate Tchebycheff inequality. *Ann. Math. Statist.*, **29**, 226-234. [61]
- PEEK, R. L. Jr. (1933). Some new theorems on limits of variation. *Bulletin Amer. Math. Soc.*, **39**, 953-959. [10, 92]
- ROMANOVSKI, V. I. (1940). On inductive conclusions in statistics. *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS (N.S.)*, **27**, 419-421. [90]
- ROYDEN, H. L. (1953). Bounds on a distribution function when its first  $n$  moments are given. *Ann. Math. Statist.*, **24**, 361-376. [43]
- SAVAGE, I. R. (1953). Bibliography of non-parametric statistics and related topics. *J. Amer. Statist. Assoc.*, **48**, 844-906. [94]
- SAVAGE, I. R. (1961). Probability inequalities of the Tchebycheff type. *J. of Res. of Nat. Bureau of Standards*, **65B**, 211-222. [87, 94]
- SELBERG, H. L. (1940). Zwei Ungleichungen zur Ergänzung des Tchebycheffschen Lemmas. *Skandinavisk Aktuarietidskrift*, **23**, 121-125. [22]
- SHOHAT, J. A. and TAMARKIN, J. D. (1943). *The problem of moments*. Amer. Math. Society. [13, 15, 48]
- TAKANO, K. (1950). A remark on Berry's paper. *Kōkyūroku of Inst. Math. Statist.*, **5**. (In Japanese.) [82]
- VON MISES, R. (1938). Sur une inégalité pour les moments d'une distribution quasi-convexe. *Bulletin des Sci. Math.* (2), **62**, 68-71. [50, 92]
- VON MISES, R. (1939). The limits of a distribution function if two expected values are given. *Ann. Math. Statist.*, **10**, 99-104. [30]
- WALD, A. (1939). Limits of a distribution function determined by absolute moments and inequalities satisfied by absolute moments. *Trans. Amer. Math. Soc.*, **46**, 280-306. [27]
- WHITTLE, P. (1958a). Continuous generalizations of Tchebycheff's inequality. *Teor. Veroyatnost. i ee Primenen.*, **3**, 386-394. [62, 63, 87]
- WHITTLE, P. (1958b). A multivariate generalization of Tchebichev's inequality. *Quart. J. of Math.* (2), **9**, 232-240. [61]
- WINSTEN, C. B. (1946). Inequalities in terms of mean range. *Biometrika*, **33**, 283-295. [36, 55]



封底

